



# SAGEO 2018

Spatial Analysis and GEOmatics

## Actes de Conférence

Maguelonne Teisseire,  
Mathieu Roche (Eds.)

6 au 9 novembre 2018

Montpellier, France



*Conférence internationale francophone*

## **Spatial Analysis and GEomatics (SAGEO'2018)**

6 au 9 novembre 2018, Montpellier, France

### **Éditorial**

Sous l'égide du **GdR MAGIS du CNRS**, la conférence internationale francophone SAGEO est un événement annuel majeur dans le **paysage de la Géomatique, de l'Analyse Spatiale et des Sciences de l'Information Géographique**.

Ses objectifs sont de présenter des travaux de recherche récents et de réunir des chercheurs de disciplines diverses (informatique, géographie, archéologie, urbanisme, sciences de l'environnement, agronomie, humanités spatiales, etc.).

Il s'agit également de promouvoir les échanges entre les différents acteurs du monde de la géomatique (chercheurs et enseignants-chercheurs, ingénieurs, gestionnaires, aménageurs, entrepreneurs...) qui œuvrent au développement et à la valorisation des recherches au sein et auprès des organismes publics, des entreprises et des collectivités territoriales.

Les articles soumis ont été évalués à travers 57 rapports d'évaluation réalisés par le Comité de Programme de SAGEO'2018 que nous remercions. L'édition 2018 de SAGEO se veut un forum ouvert à une majorité d'acteurs de la communauté géomaticienne francophone. Dans ce contexte, nous avons souhaité proposer divers types de communications qui mettent en relief la richesse et la complémentarité des travaux et des applications :

- 3 exposés invités traitant des différentes dimensions de la géomatique :
  - **Hélène Rey Valette** – *Economie* – CEE-M (Centre d'Economie de l'Environnement de Montpellier), Montpellier, France
  - **Bruno Martins** – *Traitement Automatique du Langage Naturel* – Department of Computer Science and Engineering of Instituto Superior Técnico in the University of Lisbon, Portugal
  - **Leif Isaksen** – *Humanités Numériques* – Dept. of Classics & Ancient History, University of Exeter, Grande Bretagne



- 1 tutoriel intitulé *De la Représentation de l'Espace et du Temps au sein des SIG – Mythe ou Réalité ?* proposé par **Christophe Claramunt** (Institut de Recherche de l'École navale);
- 9 articles longs;
- 8 articles courts;
- 3 articles courts présentant des démonstrations logicielles;
- 4 résumés étendus présentant des communications internationales déjà publiées.

Par ailleurs, 8 ateliers ont été sélectionnés afin d'animer les différentes communautés autour de la géomatique. Nous remercions les responsables des ateliers qui participent activement au foisonnement scientifique de ce domaine pluridisciplinaire.

Enfin, nous remercions très chaleureusement les sponsors (GDR MAGIS, Cirad, Irstea, AgroParisTech, la Région Occitanie, MUSE, #digitAg), Annie Huguet qui a piloté de multiples aspects pratiques de SAGEO'2018 et l'ensemble des membres du comité d'organisation issus de quatre unités de recherche (TETIS, ESPACE-Dev, Praxiling, IRIT) pour le travail réalisé avec entrain, enthousiasme et professionnalisme.

#### Responsables du Comité Scientifique de SAGEO'2018

*Maguelonne Teisseire, Irstea, TETIS, Montpellier*

*Mathieu Roche, Cirad, TETIS, Montpellier*

*Sylvie Servigne, INSA, LIRIS, Lyon*

*Didier Josselin, CNRS, ESPACE, Avignon*



## Comité d'Organisation

- Françoise Boissier (Irstea)
- Hugo Deleglise (TETIS)
- Jean-Christophe Desconnets (ESPACE-DEV)
- Colette Fatou (Irstea)
- Jacques Fize (TETIS)
- Sophie Fortuno (TETIS)
- Francesca Frontini (PRAXILING)
- Carmen Gervet (ESPACE-DEV)
- Annie Huguet (TETIS) - co-resp
- Dino Ienco (TETIS)
- Lynda Khiali (TETIS)
- Agnès Lherbet (TETIS)
- Josiane Mothe (IRIT)
- Jérôme Pasquet (TETIS)
- Mathieu Roche (TETIS)- co-resp
- Lucile Sautot (TETIS)
- Maguelonne Teisseire (TETIS)- co-resp
- Sarah Valentin (TETIS & ASTRE)
- Christiane Weber (TETIS)

## Comité de Programme

- Julien Andrieu
- Francois Bavaud
- Elise Beck
- Sandro Bimonte
- Alain Bouju
- Sylvain Bouveret
- Carmen Brando
- Jean-Pierre Chanet
- Jean-Pierre Chéry
- Sidonie Christophe
- Christophe Claramunt
- Paule-Annick Davoine
- Sophie de Ruffray
- Géraldine Del Mondo
- Eric Desjardin
- Rodolphe Devillers
- Thomas Devogele
- Karine Emsellem
- Laurent Etienne
- Mauro Gaio
- Gilles Gesquiere
- Jean-François Girres
- Françoise Gourmelon
- Eric Guilbert
- Jean-Christophe Desconnets
- Sylvie Lardon
- Florence Le Ber
- Marion Le Texier
- Didier Leibovici
- Therese Libourel
- Thomas Louail

- Malika Madelin
- Michel Mainguenaud
- Isabelle Mougenot
- Sebastien Mustiere
- Francois Pinet
- Sebastien Re-Coyrehourcq
- Mathieu Roche
- Christian Sallaberry
- Lucile Sautot
- Florence Sedes
- Damien Serre
- Myriam Servières
- Sylvie Servigne
- Patrick Taillandier
- Maguelonne Teisseire
- Guillaume Touya
- Pierrick Tranouez
- Marlene Villanova-Oliver
- Karine Zeitouni

## Liste des ateliers de SAGEO'2018

- **Atelier 1 : EXtraction de Connaissances à partir de donnÉEs Spatialisées (EXCES)**

*Organisateurs :*

- Christian Sallaberry, LIUPPA, Pau
- Eric Kergosien, GERiCO, Lille

- **Atelier 2 : Incertitude : des données aux modèles en Géomatique** (Atelier de l'Action Prospective du GDR MAGIS)

*Organisateurs :*

- Mireille Batton-Hubert, École nationale supérieure des Mines de Saint-Etienne
- François Pinet, IRSTEA, Clermont-Ferrand
- Eric Desjardin, CReSTIC, Université de Reims Champagne-Ardenne

- **Atelier 3 : Vers un géo-captage de masse – Capacités, limites et perspectives**

*Organisateurs :*

- Cyril Ray, IRENav, Ecole Navale.
- Karine Zeitouni, DAVID, Université de Versailles Saint-Quentin
- Mohamed Chachoua, EIVP, Ecole d'Ingénieurs de la Ville de Paris
- Rémy Thibaud, IRENav, Ecole Navale
- Zoubida Kedad, DAVID, Université de Versailles Saint-Quentin
- Yehia Taher, DAVID, Université de Versailles Saint-Quentin

- **Atelier 4 : Risques Spatialisés : Données, Modèles et Simulations**

*Organisateurs :*

- Eric Daudé, UMR IDEES, Université de Rouen
- Géraldine Del Mondo, LITIS, INSA Rouen
- Pierrick Tranouez, LITIS, Université de Rouen

- **Atelier 5 : Partage, mutualisation et aide à la compréhension de phénomènes spatio-temporels : quels liens et apports entre observatoire, infrastructure de recherche, infrastructure de données ?** (Atelier de l'Action Prospective du GDR MAGIS)

*Organisateurs :*

- Maud Loireau, Espace-DEV
- André Miralles, TETIS

- **Atelier 6 : HumaNS'2018 : Humanités numériques spatialisées**

*Organisateurs :*

- *Carmen Brando, École des hautes études en sciences sociales, EHESS Paris*
- *Francesca Frontini, Praxiling UMR 5267 CNRS, UPVM3*
- *Mathieu Roche, TETIS, Cirad*

- **Atelier 7 : Prédiction de la population et de zones habitées avec de la fusion de données d'observation de la Terre**

*Organisateurs :*

- *Michel Benet, Diginove*
- *Pierre-François Besson, Diginove*
- *Josiane Mothe, IRIT, Université de Toulouse II – Jean-Jaurès*
- *Laure Vacquie, IRIT, Université de Toulouse III – Paul Sabatier*
- *Zia Ullah, IRIT, Université de Toulouse III – Paul Sabatier*

- **Atelier 8 : Deep Learning pour la télédétection**

*Organisateurs :*

- *Jonathan Weber, IRIMAS, Université de Haute-Alsace*
- *Camille Kurtz, LIPADE, Université Paris Descartes*
- *Germain Forestier, IRIMAS, Université de Haute-Alsace*
- *Pierre Gançarski, ICUBE, Université de Strasbourg*

# Sommaire

## Session 1 - Texte

- M. Ferrand, M. El-Bèze, D. Josselin, S. Balossino and G. Castelnovo. **Traitemment informatique des entités nommées dans un corpus de données foncières en latin : méthodologie et application**.....p.9
- A.-S. Onody and E. Kergosien. **Etude de l'influence de l'environnement spatial des bibliothèques de Lille sur leur fréquentation**.....p.25
- A. Medad, M. Gaio and S. Mustière. **Appariement automatique de données hétérogènes : textes, traces GPS et ressources géographiques**.....p.41

## Session 2 - Imagerie satellitaire et végétation

- J. Dorne, N. Aussenac-Gilles, C. Comparot, R. Hugues, J.-G. Planes and C. Trojahn. **Une approche sémantique pour représenter l'indice de végétation d'images Sentinel-2 et son évolution**.....p.49
- A.-S. Onody, F. Bah, G. Duval and P. Dusseux. **Estimation de la hauteur de la végétation à partir des ombres portées d'images à Très Haute Résolution Spatiale**.....p.55

## Session 3 - Données spatiales, volontaires

- Q. Thy Truong, G. Touya and C. de Runz. **Le vandalisme dans l'information géographique volontaire : apprendre pour mieux détecter ?**..... p.61
- A. Sakka, S. Bimonte, L. Sautot, G. Camilleri, P. Zaraté and A. Besnard. **Résolution collaborative des conflits des besoins d'analyse OLAP Spatial des données issues des observatoires citoyens**.....p.77
- J. Mothe. **FabSpace 2.0 : Une innovation pédagogique qui met les géodonnées au cœur des apprentissages**.....p.83

## Session 4 - Données Géographiques sur le Web

- J. Desbonnet, F. Gourmelon and C. Claramunt. **Analyser la structure hypertextuelle des dispositifs de mise à disposition de données géographiques sur le Web**.....p.89
- M. Noucher, F. Gourmelon, G. Le Campion, A. Maulpoix, J. Pierson, O. Pissot and M. Rouan. **Vers un observatoire des données géographiques du Web. Expérimentation d'une analyse diachronique des infrastructures de données géographiques**.....p.104

## Session 5 - Modélisation

- M. Bunel, A.-M. Olteanu-Raimond and C. Duchêne. **Référencement spatial indirect : modélisation à base de relations et d'objets spatiaux vagues**.....p.118
- R. Aguejdad. **Méthodes de validation spatiale en modélisation des changements d'occupation et d'usage des sols : le cas des modèles de potentiel de transition**.....p.134

- R. Marti, T. Catry, M. Demarchi, G. L'Ambert, M. Cros, J.-S. Dehecq, N. Des-say, M. Mangeas, E. Roux, V. Herbreteau and A. Tran. **Modélisation spatiale fine de la distribution de moustiques *Aedes albopictus* par forçage météorologique et intégration de données environnementales et in situ**.....p.136
- C. Prudhomme, A. Roxin, C. Cruz, F. Boochs. **Vers la conception d'une ac-tion de réponse pour la gestion de catastrophe, utilisant la modélisati-on de connaissances**.....p.138
- R. Balzarini and F. Jambon. **De la carte au ciel : une approche empirique pour l'étude des stratégies visuelles de pilotes experts**.....p.140

## **Session 6 - Plateformes, Logiciels et Démonstrations**

- M. Sabrine Halilali, E. Gouardères, F. Devin and M. Gaio. **Plateforme lo-gicielle pour l'intégration et la composition de services géospati-aux**.....p.143
- V. Bonnal and M. Cousseau. **[Démo] Application mobile Nos pollutions - Le crowdsourcing au service de l'environnement pour caractériser les pollutions visibles**.....p.150
- A. Fonteles, M.-N. Bessagnet, A. Le Parc-Lacayrelle and C. Sallaberry. **[Démo] Un environnement pour la valorisation de données patrimoniales hétérogènes**.....p.156
- M. Viry, R. Ysebaert and M. Guérois. **[Démo] Regioviz, un outil de visualisation et de comparaison régionale**.....p.162

## **Session 7 - Humanités spatiales**

- J. Crevant, M.-H. de Sède-Marceau and S. François. **Pertinence des SI ter-ritoriaux pour la scénarisation énergétique : Démonstration par l'exemple de l'exploitation du SI OPTTEER appliqué à l'éco-mobilité**.....p.168
- G. Poujol, T. Joliveau, J. Imbernon, and F. Lançon. **Vers des cartes de flux spatialement explicites : la route du niébé ouest africain**.....p.178

## **Session 8 - Utilisateurs**

- C. Heitz, S. Marc-Zwecker, R. Mathis and F. Le Ber. **Étude exploratoire des représentations de coulées d'eau boueuse en Alsace : Analyse d'enquêtes psychosociales et de dessins à main levée via un logiciel dédié**.....p.184
- C. Moreau, T. Devogele and L. Etienne. **Extraction de motifs de trajec-toires sémantiques similaires**.....p.198
- S. Boularouk, D. Josselin and E. Altman. **Analyse d'images par réseaux de neurones à destination des malvoyants**.....p.214

---

# Traitement informatique des entités nommées dans un corpus de données foncières en latin : *méthodologie et application*

Margot Ferrand<sup>1</sup>, Marc El-Bèze<sup>2</sup>, Didier Josselin<sup>3</sup>, Simone  
Balossino<sup>1</sup>, Guido Castelnovo<sup>1</sup>

1. CIHAM, Université d'Avignon et des Pays de Vaucluse  
74 rue Louis Pasteur, 84029 Avignon, France  
[margot.ferrand@univ-avignon.fr](mailto:margot.ferrand@univ-avignon.fr)

2. LIA, Université d'Avignon et des Pays de Vaucluse

3. UMR ESPACE 73000, CNRS

---

**RÉSUMÉ.** Une approche interdisciplinaire, basée sur l'analyse de données écrites, iconographiques et archéologiques, utilisant des méthodes et des outils empruntés aussi bien aux sciences humaines et sociales qu'à la géomatique et à l'informatique, peut aujourd'hui permettre de reconstruire la typologie et la topologie des espaces urbains médiévaux dans un cadre géo-historique large. Notre recherche se focalise sur Avignon entre le XIII<sup>e</sup> et le XV<sup>e</sup> siècle. Elle est avant tout fondée sur une analyse de sources textuelles homogènes, écrites en latin, en s'appuyant sur la création d'un programme informatique de reconnaissance d'entités nommées. Ces entités, associées à d'autres types de sources, forment un corpus qui se réfère à des données essentiellement localisables. C'est pourquoi nous avons choisi de concevoir un modèle et une base de données géographiques. Ce Système d'Information Géographique permettra de confronter une grande variété de données, de les interroger et les géovisualiser, pour ainsi analyser les évolutions et les persistances de l'espace urbain avignonnais à cette époque médiévale.

---

**ABSTRACT.** An interdisciplinary approach based on the analysis of written, iconographic and archeological data using tools and methods derived from the humanities and social sciences as well, such as geomatics and information sciences, enables to reconstruct the typology and the topology of medieval urban spaces in a broad geo-historical framework. Our research focuses on Avignon between the 13th and 15th centuries and is primarily based on an analysis of homogeneous textual sources written in Latin, using a named entities recognition program. These entities, combined with other types of sources, form a corpus referring to data which have, most of the time, a location. This is why we choose to formalize a geographical data model and a data base, to build a Geographical Information System. This tool allows us to compare, query and geovisualize a wide variety of data in order to analyze the urban space evolution and structure in Avignon during this middle age historical period.

---

**MOTS-CLÉS :** Interdisciplinarité, espace urbain médiéval, entités nommées, SIG.

---

**KEYWORDS:** Interdisciplinarity, medieval urban space, named entities, GIS.

---



## 1. Introduction

*« La ville concentre de multiples intérêts contradictoires – qu'ils soient publics ou privés – et de pouvoirs tout aussi divers et souvent opposés. (...) Plus que partout ailleurs, l'espace y constitue un objet de convoitise ; plus que partout ailleurs, il est mesuré, décrit, délimité...Il s'agit d'enregistrer sur le papier ou d'inscrire par la matérialité du bâti les traces des multiples intérêts, ou encore, de faire valoir des droits. C'est dans les villes que l'espace connaît ses découpages les plus denses, c'est dans les villes aussi que les enjeux relatifs à sa maîtrise sont les plus forts. » (Arnaud, 2008).*

Depuis la seconde moitié du XX<sup>e</sup> siècle, l'histoire urbaine connaît un regain d'intérêt considérable (Mumford, 1961). La ville qui était auparavant un support ou un « cadre d'étude dont la finalité était ailleurs » (Lepetit, 1980) devient progressivement un objet d'étude à part entière. La question de la construction de l'urbain, de la « fabrique urbaine » (Noizet, 2013), est alors au centre de nombreuses recherches qui se focalisent sur la longue durée, avec une triple dimension spatiale, sociale, temporelle. Qu'il s'agisse des travaux de Bernard Lepetit, ou des dernières recherches sur les évolutions spatiales d'aires géographiques précises – Toulouse, Tours, Paris, etc. –, ces études sont inévitablement empreintes de problématiques géographiques, urbanistiques, sociologiques ou encore économiques et linguistiques et plus récemment de questionnements et de méthodes géomatiques et informatiques.

L'étude des évolutions urbaines sur une période relativement longue, antérieure à l'Ancien-Régime, nécessite l'analyse et le croisement de sources importantes aux typologies particulièrement variées : documentation textuelle, sources iconographiques, archéologiques, et cartographiques. Et si le chercheur dispose aujourd'hui d'un nombre conséquent de sources écrites éditées, numérisées et parfois déjà « océrisées », la quantité de données à traiter rend ce travail particulièrement fastidieux, voire délicat à appréhender, par une méthode historique traditionnelle. Mais, selon la documentation choisie, un prétraitement intelligent des données peut aujourd'hui être envisagé, à travers l'utilisation de l'informatique. Rappelons que l'utilisation de la machine comme outil de traitement des données n'est pas nouvelle, le colloque « Informatique et histoire médiévale » tenu à Rome en 1975 posait déjà les jalons d'un recours à la machine par le médiéviste dans certains cas spécifiques (Herlihy et Klapisch, 1978).

Ainsi, nous avons choisi de développer une approche par l'informatique pour identifier, structurer et relier entre elles, de manière automatique, les données textuelles qui nous intéressent. Cet outil permet d'appréhender au mieux la construction urbaine par l'étude de l'anthroponymie, des liens sociaux, de l'évolution des toponymes, du parcellaire et de l'organisation spatiale. Ce programme est construit pour le traitement d'une documentation foncière spécifique. Dans l'attente des cadastres – qui, à Avignon, n'arrivent qu'en 1598 –, il s'agit

essentiellement d'enquêtes administratives, de terriers et de censiers<sup>1</sup>. Il permet de traiter des corpus de taille importante, de comparer rapidement les documents et les informations qu'ils renferment grâce à l'extraction quasi automatique des entités, à leur indexation systématique dans une base de données spatiales et à leur insertion dans un Système d'Information Géographique interrogeable.

Avant d'exposer concrètement l'enjeu et le fonctionnement de ce programme de détection d'entités nommées et de comprendre dans quelle mesure il facilite l'enrichissement d'une base de données et la mise en place d'un SIG, il nous faut définir avec quel corpus de sources il a été construit, car ce sont avant tout les séries documentaires retenues pour l'étude qui en ont permis la confection. Il est donc primordial de revenir sur les sources en elles-mêmes, sur leurs typologies et sur leur contexte de production, pour pouvoir ensuite exposer le processus de construction de cet outil, les concepts sur lesquels reposent le prototype, et enfin apporter quelques premiers résultats d'analyse.

## **2. Étudier le processus évolutif des villes à partir de la documentation écrite**

Pour étudier les dynamiques urbaines médiévales à une époque pré-cadastrale, l'historien dispose d'un nombre imposant de sources écrites parmi lesquelles les données spatiales ne manquent pas. Mais à l'instar du cadastre florentin de 1427 ou du premier cadastre du Comtat-Venaissin de 1414 réalisé de manière systématique dans toutes les communautés de la région (Zerner, 1993), les documents qui précèdent cette entreprise se concentrent essentiellement sur le domaine d'une seule seigneurie. Qu'il s'agisse des enquêtes administratives d'Alphonse de Poitiers ou des livres terriers de seigneuries de moindre envergure, ces documents sont destinés à inventorier précisément les biens fonciers qui relèvent d'un seigneur, dans un but de contrôle et d'imposition de ses terres.

La ville médiévale étant un territoire fractionné entre divers pouvoirs enchevêtrés et souvent opposés, il est indispensable de croiser et de comparer l'ensemble de cette documentation, particulièrement importante, afin d'appréhender le paysage urbain et la société dans leur globalité ainsi que dans leur évolution sur une période relativement longue.

### **2.1. Une étude de cas : la cité d'Avignon**

Prenons un exemple des plus représentatifs, celui d'Avignon au bas Moyen-Âge. La cité rhodanienne est un objet d'étude idéal, puisqu'elle connaît des changements politiques et sociaux particulièrement importants. De fait, la ville connaît de 1140 à 1251 une période de relative autonomie politique avec l'instauration d'un gouvernement communal (Balossino, 2015). Elle reprend ensuite le statut de seigneurie partagée entre le comte de Toulouse et le comte de Provence jusqu'en 1290, date à laquelle les Angevins deviennent les seuls propriétaires de la cité. Si la

---

<sup>1</sup> Terriers et censiers : au Moyen Âge et sous l'Ancien Régime, registres fonciers d'un seigneur contenant la liste de ses tenanciers et de leurs tenures avec les redevances et les services auxquels ils étaient astreints.

papauté achète la ville en 1348 à la reine Jeanne, son installation progressive sur les bords du Rhône a commencé un demi-siècle plus tôt et a ainsi participé, dès le début du XIV<sup>e</sup> siècle, au renouvellement de la population et à de multiples aménagements de l'espace urbain. La cour pontificale amorce un premier retour à Rome en 1377, mais il faut attendre 1403 pour que Benoît XIII, dernier pape avignonnais du Grand Schisme, quitte définitivement son palais. La ville est, dès lors, administrée par une légation pontificale pendant plusieurs siècles et ce, jusqu'au rattachement d'Avignon à la France à la suite de la Révolution française. Pendant tout le bas Moyen-Âge, la cité avignonnaise connaît diverses dominations et chacune d'entre elles participe à sa manière à l'évolution du paysage urbain.

## ***2.2 Richesse des sources pré-cadastrales***

La situation politique et sociale complexe de la ville en fait un objet d'étude très intéressant, mais c'est avant tout la richesse de ses sources et surtout les nombreuses éditions qui ont été réalisées depuis le XIX<sup>e</sup> siècle jusqu'à nos jours qui permettent d'envisager une étude innovante du territoire urbain médiéval.

Dans une volonté de contrôle et de domination territoriale, les différentes institutions qui se sont succédé dans la ville ont toutes pris soin de recenser les biens qui relevaient de leur autorité. En 1233, Bertrand du Pont, notaire de la commune, est chargé de dresser un inventaire des biens et des ressources fiscales de l'institution à partir d'une série d'enquêtes sur les propriétés collectives (De Maulde, 1879). À son arrivée à la tête de la Provence en 1246, Charles I<sup>er</sup> d'Anjou demande la réalisation d'une enquête sur le domaine dont il vient d'hériter par son mariage avec Béatrice, fille du comte Raymond Bérenger V. Cette enquête a un double but : établir d'une part les droits domaniaux et récupérer d'autre part tous les droits que les prédécesseurs du comte avaient exercés (Baratier, 1969). Moins de cinquante ans après la réalisation de l'enquête de Charles I<sup>er</sup> de Provence, son fils Charles II lance la seconde enquête générale de Provence achevée en 1298 (Pécourt, 2018). Enfin, la troisième enquête réalisée pour ce même territoire, de loin la plus complète, a été ordonnée par Robert d'Anjou, roi de Naples et comte de Provence, et réalisée en 1333 (Pécourt, 2005).

Si l'étude de ces enquêtes permet d'appréhender l'évolution d'une partie du territoire avignonnais entre le XIII<sup>e</sup> et le début du XIV<sup>e</sup> siècle, c'est leur croisement avec un autre type de documents qui permet d'avoir une vision plus globale du paysage urbain de la cité. En effet, pour le XIV<sup>e</sup> et le XV<sup>e</sup> siècle, nous disposons, en parallèle de ces grandes enquêtes administratives et foncières, d'un nombre important de terriers (Hayez, 1993) et de censiers. Il ne faut pas oublier que si la ville appartient à la papauté depuis 1348, son territoire n'en est pas moins fractionné. Une multitude d'institutions et de particuliers possèdent ainsi des droits dans Avignon. Le paysage urbain avignonnais est donc scindé entre le pape, l'évêque d'Avignon, les différents chapitres et institutions ecclésiastiques, ainsi que certains propriétaires indépendants qui ont acquis des cens au fil du temps.

On dénombre ainsi, pour le seul XIV<sup>e</sup> siècle, une dizaine de terriers aux archives départementales de Vaucluse (ADV), et d'innombrables reconnaissances dispersées entre les fonds. À l'instar des grandes enquêtes administratives, les terriers ont fait

l'objet d'études universitaires notamment depuis les années 1930. Au départ, il s'agit d'une initiative prise par des géographes ruralistes, puis les historiens des Annales ont également su trouver en ces documents des sources de qualité en matière d'économie et de société. Souvent transcrits, voire édités, ces documents renferment une quantité impressionnante d'informations qui permettent d'étudier l'évolution du paysage mais également du système seigneurial en lui-même, des individus, de leur pratique et de leurs liens familiaux (Brunel *et al.*, 2002).

Ces documents sont parfaitement complémentaires, et, pour étudier l'espace urbain dans sa globalité, il faut les assembler dans une base de données et recouper l'ensemble des informations qu'ils délivrent.

### 3. Modèles Conceptuels de Données à partir des points communs des sources

Si on se concentre sur Avignon, que ce soit l'inventaire de la commune de 1233, les grandes enquêtes angevines de 1251, 1298 et 1333, ou les nombreux terriers du XIV<sup>e</sup> et XV<sup>e</sup> siècles, ces sources, de prime abord peu homogènes, comportent de nombreux points communs qui permettent de réaliser des Modèles Conceptuels de Données relativement génériques (implémentés sous PostGreSQL/PostGIS) et d'étudier les textes se référant à ces entités de manière quasi automatique (application *AutoAnnot* développée sous Unix/linux).

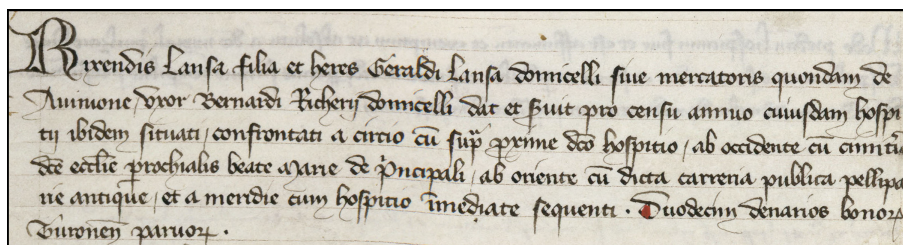


FIGURE 1. Extrait du Terrier de l'évêque d'Avignon (1366-1368), ADV, 1G10.

Ces documents, tous rédigés en latin, énumèrent les biens d'une autorité dans la ville à un moment donné (cf. Fig. 1). Selon la date à laquelle ils sont réalisés, leur commanditaire et le but de leur confection, ces enquêtes sont plus ou moins précises. Elles sont toutefois toutes construites de la même façon, puisqu'elles recensent de manière systématique les individus qui doivent payer une redevance à un seigneur spécifique pour une parcelle qu'ils détiennent dans la ville (cf. Fig. 2). Il peut s'agir de jardins, de vignes, de magasins, de maisons ou de véritables hôtels, avec cour et dépendances. Les informations délivrées par ces sources sont manifestement organisées selon le même schéma. En effet, dans tous les documents, une série de déclarations se suivent ; elles sont alors particulièrement répétitives.

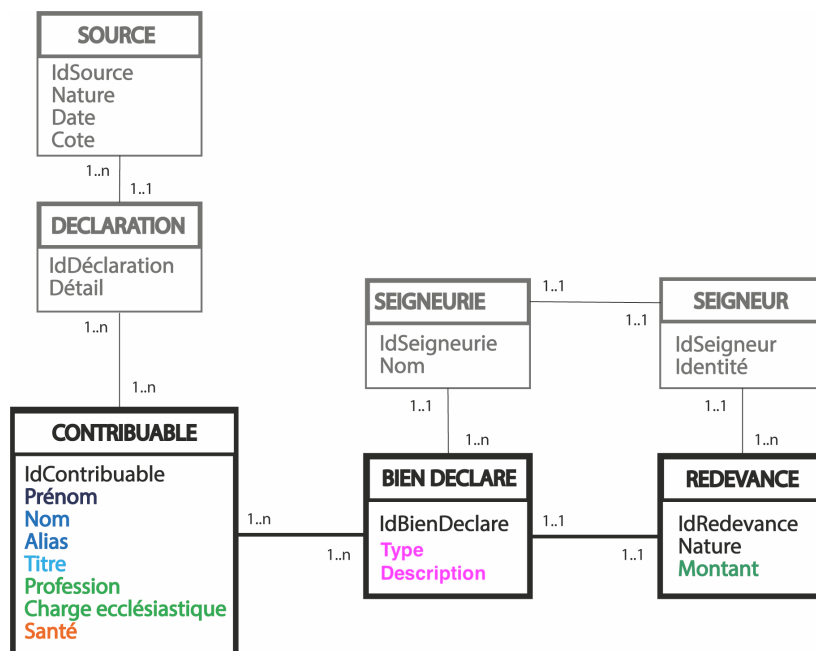


FIGURE 2. Diagramme UML - Partie centrale MCD - Les couleurs renvoient aux Entités Nommées détectées automatiquement par AutoAnnot (cf. Fig. 5)

Dans chacune de ces sources, l'identité du tenancier est donnée en premier. Elle est définie de différentes manières, peut contenir par exemple le métier de la personne, son statut social et être complétée par ses liens de parenté, de familiarité ou encore professionnels (cf. Fig. 3).

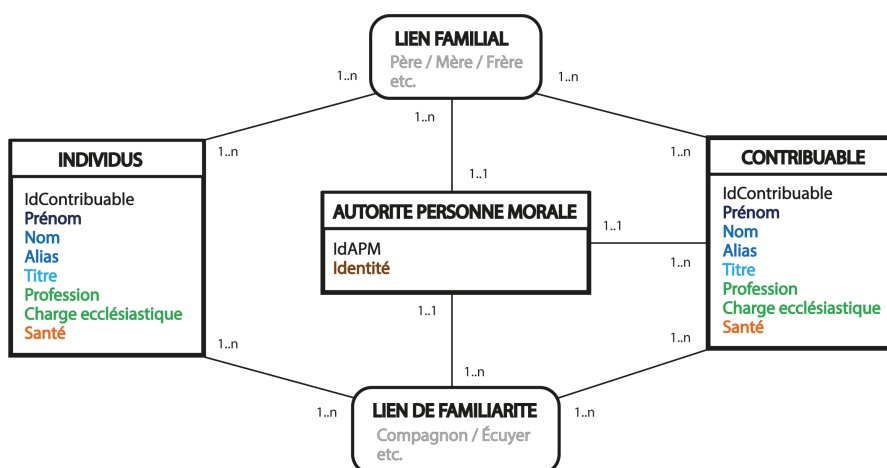


FIGURE 3. Diagramme UML - Liens entre les personnes (MCD suite)

Il est ensuite question du bien détenu par le contribuable. Le plus souvent dénommée par sa simple typologie, une description plus précise de la propriété est parfois donnée (état de conservation du bien, matériaux de construction, mesure). La localisation du bien (cf. Fig. 4) est ensuite régulièrement renseignée en fonction de ses *confronts*<sup>2</sup>, de son orientation par rapport à la rue ou encore de sa proximité avec un édifice emblématique de la cité, ce qui confère au bien une localisation relative (Hautefeuille, 2016).

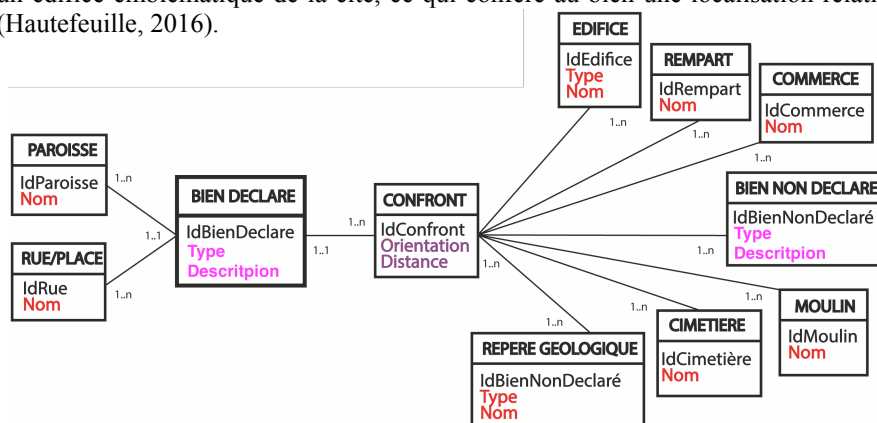


FIGURE 4. Diagramme UML simplifié - Localisation des biens déclarés (MCD fin)

Le caractère répétitif de ces différents documents, que ce soit dans la forme même des déclarations ou dans leur fond, permet de les appréhender de manière quasi automatique par un traitement informatique basé sur la détection des entités nommées. Ce traitement facilite considérablement l'analyse historique, puisqu'il permet d'alimenter automatiquement la base de données à partir de l'extraction des entités, et de croiser ainsi un nombre conséquent de grands volumes d'archives.

#### 4. La construction d'un programme de reconnaissance des entités nommées

Historiquement, le terme d'entité nommée (EN) désigne « l'ensemble des noms propres présents dans un texte, qu'il s'agisse de noms de personnes, de lieux ou d'organisation, ensemble auquel sont souvent ajoutées d'autres expressions comme les dates, les unités monétaires, les pourcentages et autres » (Ehermann, 2008). Depuis les premières recherches sur ce sujet, la définition a toutefois évolué, les entités nommées représentent aujourd'hui les données pertinentes déterminées et structurées en fonction d'une application concrète. Chaque entité nommée se réfère alors à « une entité extralinguistique ou à un référent unique » et « représente une expression linguistique autonome, elle permet à elle seule l'identification du référent » (Hatmi, 2014). Le développement des systèmes de reconnaissance d'entités nommées touche également l'information géographique, avec ses spécificités (Lesbegueries *et al.*, 2006 ; Moncla *et al.*, 2016 ; Palacio *et al.*, 2016).

<sup>2</sup> L'espace des *confronts* est un espace de la contiguïté, littéralement « confiné » (Penet, 2006).

Des travaux traitant des méthodes d'annotation d'expressions appliquées à des données spatiales et historiques existent également (Gaio *et al.*, 2012).

Mais à ce jour, et à notre connaissance il n'existe qu'un seul programme de détection d'entités nommées pour la langue latine. Il a été réalisé pour structurer et rendre interrogeable la plate-forme documentaire *Corpus Burgundiae Medii Aevi* (CBMA)<sup>3</sup>, projet dédié à l'étude de la documentation diplomatique bourguignonne du Xe au XIII<sup>e</sup> siècle. Ce programme, construit à partir d'un corpus d'entraînement annoté manuellement, ne distingue que deux entités : la première se réfère aux personnes et la seconde aux lieux (Torres Aguilar, 2017). Notre programme reconnaît quant à lui quinze entités.

Notre corpus n'ayant pas les mêmes caractéristiques et l'enjeu de notre travail n'étant pas le même, nous avons créé un nouveau programme pour classer, relier et indexer automatiquement les nombreuses informations qui nous intéressent dans la documentation foncière du bas Moyen-âge écrite en latin.

#### 4.1. Les règles du programme d'étiquetage

Le programme que nous avons développé repose sur une base de règles préalablement définies en fonction des sources. Il est construit sur un travail sémantique. Il ne s'agit pas donc d'un étiquetage syntaxique mais bien lexical, le but étant d'attribuer une classe à chaque entité, quelle que soit sa fonction dans la phrase. Un lemmatiseur a également été mis en place pour réduire la variabilité de certaines graphies, normaliser la détection et améliorer l'analyse statistique. Après avoir passé en minuscules et lemmatisé le corpus, tout en gardant trace des formes fléchies, des Cascades de Transducteurs sont appliquées (Maurel et Friburger, 2004)<sup>4</sup>. Un transducteur n'est rien d'autre qu'un automate qui réalise conjointement une analyse du texte et la production d'annotations (étiquettes) pour enrichir les documents. Dans notre cas, les étiquettes sont les EN et les couleurs qui leur sont associées (cf. Fig. 5).

Dans le cas de notre étude, et au vu de notre corpus de sources, trois classes d'entités centrales ont tout d'abord été définies de manière évidente (cf. Fig. 2 et 5) : celles concernant les individus cités dans ces enquêtes, celles se référant aux biens déclarés (TyBien) et celles qui donnent le montant de la redevance versée annuellement par chaque contribuable au seigneur (Redev). Plusieurs classes ont ensuite été ajoutées. L'une d'entre elles, divisée en sous-catégories pour plus de précision, concerne les lieux fixes de la ville (LOC) – édifices emblématiques, lieux de culte, marchés, remparts, etc. En contexte, les termes de cette classe sont généralement introduits par des adverbes de lieu, des prépositions ou des locutions prépositionnelles regroupés également dans une classe (PosRel). Cette dernière classe est elle-même départagée en sous-classes pour plus de précision : on trouve ainsi l'information sur la direction, l'orientation ou encore la relation de voisinage. Ce surcroît de précision permet une analyse géographique plus fine.

<sup>3</sup> [cbma-project.eu](http://cbma-project.eu)

<sup>4</sup> Le terme de cascades vient du fait que plusieurs automates sont appliqués de façon séquentielle lors de passes successives dont l'enchaînement sert à affiner l'étiquetage.

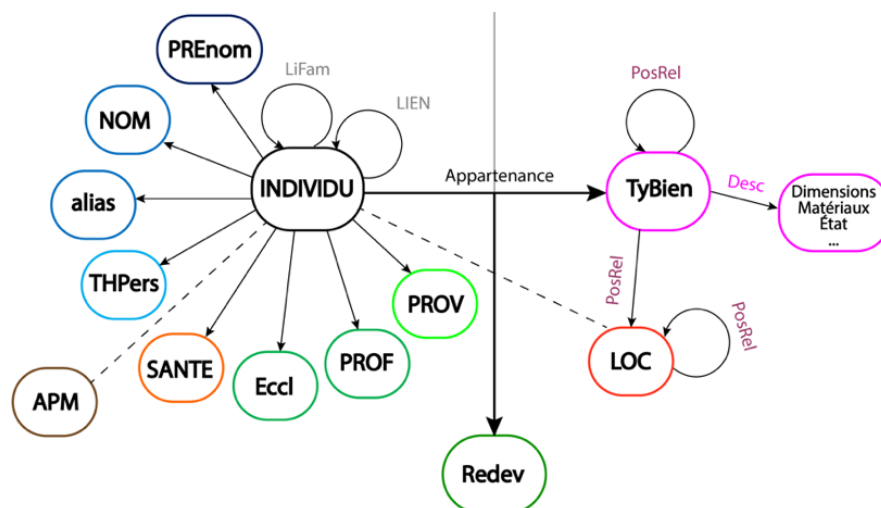


FIGURE 5. Structure d'exploration des entités et de leurs liens

Les entités se référant aux individus sont départagées en différentes classes qui permettent d'effectuer des études ciblées selon les problématiques. On trouve ainsi les entités donnant les prénoms (PREnom), les noms (NOM) et les surnoms (Alias), mais également celles se référant aux titres honorifiques (THPers) et à la provenance des personnes (PROV), celles définissant leur profession (PROF) ou leur charge ecclésiastique (Eccl) et enfin celles donnant des précisions sur leur état de santé (SANTÉ ; il s'agit le plus souvent de personnes décédées et citées dans les enquêtes). Deux entités destinées à permettre d'étudier les liens familiaux et sociaux qui unissent les individus ont également été créées (LiFam ; LIEN). Enfin, nous avons fait le choix de créer une entité spécifique (APM) définissant l'autorité dans la ville (exemple de l'évêque) ou encore une personne morale (la commune d'Avignon). Deux entités nous ont servi à définir les biens déclarés dans la ville : l'une donne la typologie du bien en question – maison, vigne, étable, etc. – et l'autre sa description quand elle est présente.

26 EN\_EN=PREnom RIXENDIS\_EN EN\_EN=NOM LANSA\_EN EN\_EN=LIFAM filia\_EN et EN\_EN=LIFAM heres\_EN  
 EN\_EN=PREnom Geraldus\_EN EN\_EN=NOM Lansa\_EN EN\_EN=THPers domicelli sive mercatoris\_EN EN\_EN=SANTÉ  
 quondam\_EN EN\_EN=PROV de Avinione\_EN EN\_EN=LIFAM uxor\_EN EN\_EN=PREnom Bernardi\_EN EN\_EN=NOM  
 Richerii\_EN EN\_EN=THPers domicelli\_EN dat et servit pro censu annuo EN\_EN=TyBien hospitii\_EN  
 EN\_EN=PosRel\_localisation ibidem situati\_EN EN\_EN=PosRel\_confront confrontati\_EN EN\_EN=PosRel\_a\_nord a circio\_EN  
 cum dicto EN\_EN=TyBien hospitio\_EN EN\_EN=PosRel\_a\_Loccident ab occidente\_EN cum EN\_EN=LOC-Cimetieres cimiterio  
 dicte ecclesie parochialis Beate Marie de Principali\_EN EN\_EN=PosRel\_a\_Lorient ab oriente\_EN cum EN\_EN=LOC-Rues carreria  
 publica Pelliparie antique\_EN et EN\_EN=PosRel\_aud a meridie\_EN cum EN\_EN=TyBien hospitio\_EN immediate sequenti  
 EN\_EN=MonRed 12 den. bon. tur. parv\_EN

FIGURE 6. Exemple d'étiquetage automatique



#### 4.2. Lien entre les entités

Le lien entre les entités peut ensuite être détecté à partir de l'étude des cooccurrences des entités dans la proximité immédiate et la définition d'un grand nombre de règles. Les entités formant l'identité d'un individu sont organisées selon des motifs récurrents ; par exemple : PREnom + NOM + Alias + PROV + THPers + PROF/Eccl. Le lien d'appartenance entre les entités définissant un individu et une propriété est ensuite détecté par l'expression « *pro censu* » ou « *dat et servit pro censu* », une variante de celle-ci. La relation familiale entre les individus et donc entre les entités s'y référant, a quant à elle été définie de différentes manières. Elle peut être de type binaire ; il s'agit de la plus courante et de la plus facile à identifier : *untel* est défini comme étant le fils de telle personne (cf. Fig. 7).

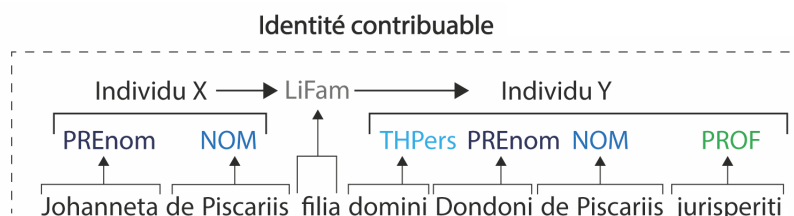


FIGURE 7. Motif le plus courant exprimant le lien familial entre deux individus

La relation peut également être ternaire (tel individu et tel autre individu sont les enfants et les héritiers de telle personne), plus généralement *Naire* orientée ou *Naire* circulaire (trois personnes citées successivement sont dites frères), et dans certains cas unaires, lorsqu'une personne est simplement définie comme l'héritière de quelqu'un, sans autre indication sur sa propre identité. Différents motifs sont ainsi reconnus. Outre les relations familiales, les liens professionnels sont établis lorsque la charge exercée est directement rattachée à un lieu (exemple des ecclésiastiques) ou qu'une personne travaille pour un individu, une autorité ou une institution (camérier du pape, cuisinier d'un cardinal). Puisqu'on intègre également la notion de provenance des individus et de localisation de leur bien, c'est un véritable réseau social complexe et enchevêtré qu'il est potentiellement possible d'établir et de géoréférencer de façon relative.

#### 4.3. Lever les ambiguïtés

Il n'est pas toujours aisé de déterminer la classe d'une entité, et les ambiguïtés linguistiques vont de pair avec tout programme de détection d'entités nommées, quel que soit son modèle. Si certaines ne peuvent être réglées par l'outil informatique et relèvent de l'analyse historique même – exemple de l'ambiguïté entre la classe des noms et celles des provenances –, d'autres peuvent être résolues par le biais de diverses règles liées à l'analyse du document.

L'ambiguïté la plus courante que nous rencontrons concerne évidemment les rapports entre les prénoms et les noms. Elle est toutefois rapidement levée à partir

d’une étude préalable des textes et du traitement informatique (cf. Fig. 8). En effet, une première analyse des documents a permis de dresser une série d’observations facilitant la désambiguïsation. Le prénom d’un individu est toujours cité avant son nom quand ce dernier est renseigné. Dans ce cas, le programme retenant toujours la séquence la plus longue, nous avons établi la règle selon laquelle si deux prénoms ou deux noms se suivent immédiatement, alors la première entité doit être étiquetée comme prénom et la seconde comme nom. Les ambiguïtés entre les entités définissant des noms et des professions sont également courantes. Dans ce cas, nous nous sommes appuyés sur le travail préalable de l’éditeur pour résoudre le problème. De fait, lorsqu’il s’agit d’un nom et non d’une profession, l’entité possède une majuscule. Le programme est donc entraîné à choisir la classe en conséquence.

<p>EN_EN=NOM Petrus_EN EN_EN=NOM Albertus_EN  EN_EN=NOM Alfantus_EN EN_EN=PROF Fusterius_EN</p>	<p>EN_EN=PREnom Petrus_EN EN_EN=NOM Albertus_EN  EN_EN=PREnom Alfantus_EN EN_EN=NOM Fusterius_EN</p>
---	--

FIGURE 8. Exemple d’étiquetage avant et après la désambiguïsation

Les titres honorifiques des individus peuvent précéder leurs noms, tout comme les suivre. Il est d’ailleurs possible qu’une personne possède deux titres. L’un précède alors l’identité et l’autre la suit directement : **Dominus** Berardus Rascatii **miles** cité dans le terrier d’Anglic Grimoard. Dans ce cas, le programme *AutoAnnot* reconnaît le motif **Titre honorifique** + *Prénom* + *Nom* + **Titre honorifique**. La redevance que doit payer le contribuable est quant à elle systématiquement citée à la fin de la déclaration.

La majeure partie des ambiguïtés entre classes, résolue ou non par l’outil informatique, ne doit en aucun cas être dissociée du contexte historique et géographique de l’étude. En effet, ces ambiguïtés révèlent régulièrement les évolutions linguistiques et sociales que l’on rencontre d’un document à un autre, d’une époque à une autre.

L’outil *AutoAnnot* a été testé sur une documentation étrangère à notre corpus de source pour évaluer sa robustesse. Les premiers résultats sont particulièrement satisfaisants, avec plus de 80% de reconnaissances des EN concernant les individus et les lieux. Toutefois, comme tous les modèles de détection d’entités nommées il s’agit d’un modèle *ad hoc*. De plus, sa couverture n’est pas encore totalement complète, il doit donc être constamment enrichi pour gagner en performance.

## 5. Perspectives d’analyses (quantitatives / qualitatives)

### 5.1 Un besoin accru d’informations historiques et spatialisées

Si notre corpus commence au début du XIII<sup>e</sup> siècle et que le nom unique est déjà en grande partie remplacé par le nom à deux éléments, certaines personnes ou groupes sociaux échappent à cette généralisation (Bourin, 1989-2008). Il n’est en effet pas rare de rencontrer dans les premières enquêtes du XIII<sup>e</sup> siècle des individus

cités par leur seul prénom. Dans les derniers documents que nous étudions, seuls les clercs et les femmes semblent encore être désignés ainsi. Certaines exceptions nous poussent toutefois à nuancer notre propos. En effet, si les femmes sont habituellement désignées par leur prénom et leur lien familial, le plus souvent paternel ou marital, les nobles et les femmes issues des parentés les plus aisées font exception et sont désignés par leur prénom et leur nom.

Une phase de transition pendant laquelle le « nom » ne peut pas encore être réellement défini comme le nom de famille – dans le sens actuel de son appellation, sa transmission n'étant pas encore effective –, peut également être décelée à partir de l'outil informatique. L'extraction des entités nommées permet en effet d'analyser la formation même de ces noms par le glissement de certaines entités d'une classe à une autre. Ainsi, il est courant de rencontrer le nom du père comme second nom : *Petrus filius Martini* devient par exemple, selon les documents, leur commanditaire et la date à laquelle ils sont réalisés, *Petrus Martini*. Il est parfois question d'autres relations familiales, de surnom (généralement précédé par le terme « alias », c'est ce qui permet au programme de le détecter), ou encore de désignations complémentaires d'ordre professionnel ou social. De nombreux locatifs font également office de second nom, c'est ici qu'une première dimension spatiale intervient, il peut alors s'agir d'un lieu de seigneurie, de domaine, ou de résidence d'origine de la famille, par exemple *Johannes de Sancto Genesio* ou *Ugo de Rognonascio* cités dans l'enquête de 1255.

Outre l'anthroponymie, l'évolution de la toponymie peut également être pré-analysée à partir du traitement automatique des entités nommées. L'étude des extractions permet effectivement de déceler la formation progressive des noms de rues et de places, ainsi que leur évolution. Dans l'inventaire de la commune de 1233 par exemple, une seule rue est désignée par un nom, *Fusteria Nova*, et celui-ci est directement lié à l'activité commerciale du quartier, celle des fustiers. Au fil des enquêtes, les noms apparaissent et se figent progressivement. Ainsi, dans le terrier de l'évêque avignonnais de 1366-1368, on trouve une vingtaine de noms de rues et de places. Si certains sont de plus en plus précis avec le temps et font alors référence à un établissement religieux, à une famille ou à un personnage illustre, d'autres sont encore désignés par le trajet que les rues servent à parcourir (Hayez, 1993).

Enfin, la détection des EN nous permet d'appréhender l'évolution de l'espace urbain et la manière dont il est décrit à partir de l'étude de la typologie des biens déclarés sur un même territoire – souvent une paroisse ou un quartier – à des dates différentes ou par des commanditaires différents. Les terrains non bâtis, en latin *locale*, ou encore les jardins, sont très présents dans l'intra-muros du XIII<sup>e</sup> siècle, et disparaissent au profit de construction de véritables demeures avec l'installation de la papauté dans la ville au début du XIV<sup>e</sup> siècle (Balossino, 2013). L'outil de détection d'entités nommées permet ainsi de déceler rapidement et d'étudier en conséquence les évolutions dans la manière de décrire l'individu, son territoire et ses biens. Il est également possible de faire des études sur l'enrichissement de certains quartiers ou de certaines demeures à travers l'étude des redevances.

Au fil des enquêtes, on décèle un besoin accru d'information et de détails sur les hommes, leur situation, leurs biens et leur localisation. Alors que les individus sont cités dans les premières enquêtes par leur seul prénom ou par leur « double nom »,

ils sont très généralement désignés au XIV<sup>e</sup> siècle par leur prénom, leur nom, leur surnom, leur titre honorifique, leur métier mais également leur provenance et leurs liens familiaux. À une période où la démographie d'Avignon ne cesse de varier avec la venue de courtisans, de marchands, de banquiers et de voyageurs de tous horizons, les autorités démontrent un besoin grandissant de recenser la population et bien sûr de l'imposer. Le contrôle de l'espace urbain et de ceux qui le façonnent est alors un enjeu primordial que la multiplication des sources foncières et des informations qu'elles renferment permet de démontrer.

## 5.2. Interrogation et visualisation des données historiques socio-spatiales

Afin de reconstruire la typologie de l'espace urbain médiéval, l'ensemble des documents écrits est indexé dans la base de données à partir de l'extraction automatique des entités nommées. Ces sources doivent évidemment être croisées avec une documentation bien plus large, iconographique, archéologique et cartographique. C'est la confrontation de ces données dans un Système d'Information Géographique qui permet de visualiser concrètement les évolutions socio-spatiales, de dégager l'impact éventuel des gouvernements sur la politique urbaine ainsi que l'influence de certains individus, notamment les cardinaux, sur la construction d'un quartier. L'analyse "régressive" de l'espace urbain fondée sur l'étude des cartes historiques (cf. Fig.9), l'analyse archéologique du bâti et le dépouillement systématique de la documentation écrite se référant à des données localisables permettent en effet de reconstruire le parcellaire médiéval, de dresser l'état de certains quartiers à un moment donné et d'appréhender leur formation et leur évolution à partir des transformations sociétales et politiques que connaît la cité.



FIGURE 9. Île 134 : au centre cadastre actuel, à gauche plan d'alignement des façades (1854), à droite gravure d'Avignon (M. Merian 1635) et cadastre napoléonien.

À Avignon, l'explosion démographique induite par l'installation de la papauté à partir de 1309 oblige les pontifes à prendre des mesures radicales pour loger l'ensemble de leur cour : des listes d'immeubles réquisitionnés et assignés sont dressées (Pansier, 1926). Entourés de leur *familia*, les cardinaux (cf. Fig. 10) se suffisent rarement des quelques maisons qui leur sont livrées et en rachètent progressivement d'autres. À l'instar de Guillaume de la Jugie et de sa cour, l'installation des cardinaux s'étend bien souvent au-delà de leurs simples propriétés et définit généralement de véritables quartiers délimités par des *cancels* (sorte de barrières). C'est le croisement des différentes sources foncières qui permet alors de déceler l'emprise réelle de ces individus sur le territoire urbain.

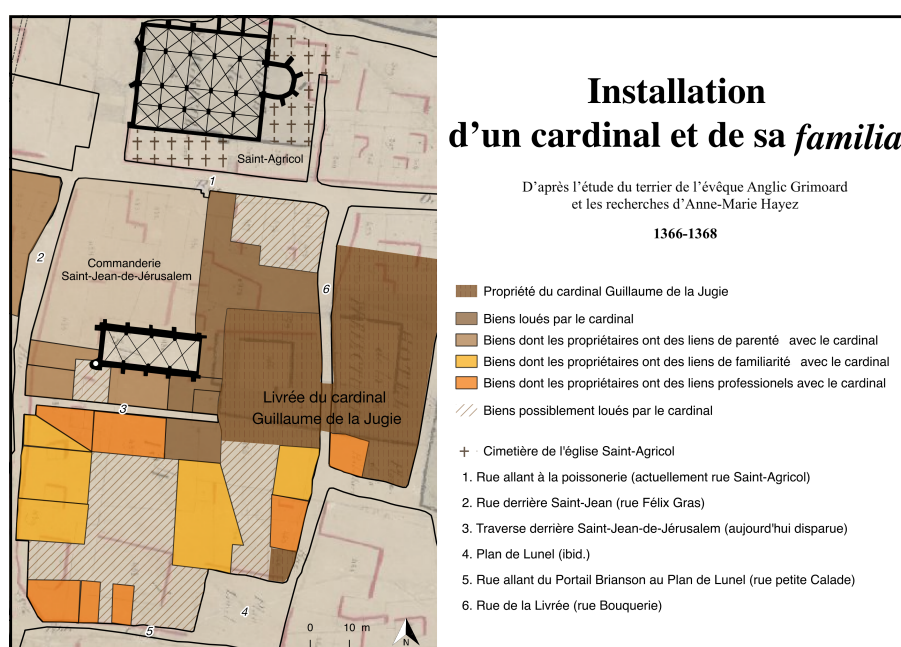


FIGURE 10 – Livrée cardinalice de Guillaume de la Jugie v. 1366-1368

L'arrivée des cardinaux et de leurs entourages modifie ainsi en profondeur l'habitat avignonnais tout en contribuant au renouvellement de la composition même d'une société urbaine, à travers le réseau socio-spatial qu'ils créent et que l'on peut observer et interroger via le Système d'Information Géographique dédié à cet effet.

## 6. Conclusion

Afin d'étudier le processus évolutif des villes au Moyen-Âge à partir des sources pré-cadastrales, une quantité impressionnante d'informations peut être regroupée. Pour optimiser l'analyse de notre corpus textuel, nous avons construit

un programme de reconnaissance des entités nommées. Ces entités ont été préalablement définies pour faciliter le classement de nos données, leur interrogation et leur visualisation dans un SIG. C'est en effet l'indexation systématique de ces données et leur croisement avec une documentation hétérogène qui permet aujourd'hui de reconstruire la typologie de l'espace urbain médiéval et de mettre en exergue les facteurs de son développement.

L'enjeu de l'outil *AutoAnnot* est double : il permet d'effectuer une étude quantitative des données par le traitement et le croisement d'un grand nombre d'entités repérées et une étude qualitative de ces mêmes données, avec la mise en perspective des indications qu'elles donnent sur l'évolution des pratiques sociales, dans le temps et dans l'espace. Le programme a été en premier lieu développé et testé sur une documentation foncière avignonnaise rédigée en latin entre le début du XIII<sup>e</sup> au XV<sup>e</sup> siècle. Sa couverture doit à présent faire l'objet d'un enrichissement automatique de vocabulaire pour pouvoir être utilisé sur une documentation bien plus vaste tant du point de vue géographique que chronologique. Ce travail s'inscrit dans une perspective plus large de thèse, financée par la FR Agorantic, sur l'impact socio-spatial de l'arrivée de la papauté à Avignon au Moyen-Âge.

## Bibliographie

- Arnaud J.-L. (2008). *Analyse spatiale, cartographie et histoire urbaine*, Parenthèses/MMSH, Marseille.
- Balossino S. (2013). L'enquête de Leopardo da Foligno dans la viguerie d'Avignon, T. Pécourt (dir.) *L'enquête générale de Leopardo da Foligno en Provence Occidentale*, CTHS, Paris.
- Balossino S. (2015). *I podestà sulle sponde del Rodano, Arles e Avignone nei secoli XII e XIII*, Viella, Rome.
- Baratier É. (1969). *Enquêtes sur les droits et revenus de Charles I<sup>er</sup> d'Anjou en Provence (1252 et 1278)*, Paris.
- Bourin M. (1989-2008). *Genèse médiévale de l'anthroponymie moderne : étude d'anthroponymie médiévale*, Bibliothèque de l'école des chartes, Tours.
- Brunel G., Guyotjeannin O., Moriceau J.-M. (éd.) (2002). Terriers et plan-terriers du XIII<sup>e</sup> au XVIII<sup>e</sup> siècle. *Actes du colloque de Paris 23-25 septembre 1998*, Rennes.
- De Maulde R. (1879). *Coutumes et règlements de la République d'Avignon au treizième siècle*, Paris.
- Ehermann M. (2008). *Les Entités Nommées, de la linguistique au TAL : Statut théorique et méthode de désambiguïsation*. Thèse en Informatique et langage, Université Paris Diderot.
- Friburger N., Maurel D. (2004). Finite-state transducer cascades to extract named entities in textes, *Theoretical Computer Science*, 313, Essex.
- Gaio M., Sallaberry C., Nguyen V. T. (2012). Typage de noms toponymiques à des fins d'indexation géographique, *Traitement automatique des Langues, ATALA*, Vol. 53.
- Hautefeuille F. (2016). *Géolocalisation des sources fiscales pré-révolutionnaires : la quadrature du cercle*, <http://journals.openedition.org/cem/13800>.

- Hatmi M. (2014). *Reconnaissance des entités nommées dans des documents multimodaux*. Thèse en Informatique, Université de Nantes.
- Hayez A.-M. (1993), *Le terrier de l'évêque Anglic Grimoard (1366-1368)*, CTHS, Paris.
- Herlihy D., Klapisch C. (1978). *Les Toscans et leurs familles, une étude du catasto florentin de 1427*, Presses de la Fondation nationale des Sciences Politiques et Édition de l'EHESS, Paris.
- Lepetit B. (1980). Histoire urbaine et espace. *Espace géographique*, vol. 9, n°1.
- Lesbegueries J., Sallaberry C., Gaio M. (2006). Associating spatial patterns to text-units for summarizing geographic information, ACM. *ACM SIGIR, Geographic Information Retrieval, Workshop*, Seattle.
- Moncla L., Gaio M., Nogueras-Iso J., Mustière S. (2016). Reconstruction of itineraries from annotated text with an informed spanning tree algorithm, *International Journal of Geographical Information Science*, vol. 30, n. 6, [doi:10.1080/13658816.2015.1108422](https://doi.org/10.1080/13658816.2015.1108422).
- Mumford L. (1961). *The city in history. Its origins, its transformations, and its prospects*, Harcourt, New York.
- Noizet H. (2007). *La fabrique de la ville. Espace et sociétés à Tours (IX<sup>e</sup>-XIII<sup>e</sup> siècles)*. Publications de la Sorbonne, Paris.
- Noizet H., Bove B., Costa L. (2013). *Paris de parcelles en pixels : Analyse géomatique de l'espace parisien médiéval et moderne*, Presses universitaires de Vincennes, Paris.
- Palacio D., Sallaberry C., Cabanac G., Hubert G. (2016). *Cadre d'évaluation de systèmes de reconnaissance d'entités nommées spatiales*. 34<sup>ème</sup> Congrès Informatique des Organisations et Systèmes d'Information et de Décision (INFORSID 2016). Grenoble.
- Pansier P. (1926). *Les palais cardinales d'Avignon aux XIV<sup>e</sup> et XV<sup>e</sup> siècle*, Roumanille, Avignon.
- Pécourt T. (dir.), Bouiron M., Butaud G., Jansen P., Venturini A. (éd.) (2008). *L'enquête générale de Leopardo da Foligno en Provence orientale (avril-juin 1333)*, CTHS, Paris.
- Pécourt T. (dir.), Bouiron M., Bufo P., Gouiran G., Lasalle J., Olivieri A., Otchakovsky-Laurens F., Rao R., Sauze É., Sibon J.(éd.) (2018). *L'Enquête générale de Charles II en Provence (1297-1299)*, Comité des Travaux Historiques et Scientifiques (CTHS), Paris.
- Penet H. (2006). Le sens des limites. Construction et perception de l'espace dans les actes de la pratique : l'exemple sicilien (XII<sup>e</sup>-XV<sup>e</sup> s.), *Actes des congrès de la SHMESP*, 37<sup>e</sup> congrès, Mulhouse.
- Torres Aguilar S. (2017). La reconnaissance des entités nommées dans les bases numériques de chartes médiévales en latin : le cas du Corpus Burgundiae Medii Aevi (X<sup>e</sup>-XIII<sup>e</sup> s.), *Médiévales*, 73, Vincennes.
- Zerner M. (1993). *Le cadastre, le pouvoir et la terre. Le comtat Venaissin pontifical au début du XV<sup>e</sup> siècle*, École française de Rome, Rome.

---

# Etude de l'influence de l'environnement spatial des bibliothèques de Lille sur leur fréquentation.

Anne-Sophie Onody<sup>1</sup>, Eric Kergosien<sup>2</sup>

1. AgroCampus Ouest

anne-sophie.onody@agrocampus-ouest.fr

2. Laboratoire GERiCO, Université de Lille - SHS

eric.kergosien@univ-lille.fr

---

*RÉSUMÉ.* Les bibliothèques françaises connaissent depuis quelques années une baisse de fréquentation. Le réseau des bibliothèques de la Métropole Européenne de Lille (MEL) n'échappe pas à ce phénomène. Dans le cadre du projet LivreEtLecture, nous menons une étude afin d'analyser l'insertion de ces établissements sur le territoire de la MEL au regard d'autres types de lieux et d'en évaluer les conséquences sur leur fréquentation. Pour cela, nous comptabilisons le nombre de chaque type d'infrastructures autour des bibliothèques pour établir des 'signatures spatiales' que nous comparons ensuite entre elles selon le taux d'emprunteurs actifs des établissements. Pour affiner le choix des données utilisées et l'échelle d'étude, des entretiens avec l'usage de cartes mentales furent menés auprès des usagers de différentes bibliothèques. Dans cet article, nous présentons les expérimentations menées sur les bibliothèques de Lille.

*ABSTRACT.* For years French libraries have experienced a decline of their attractiveness and so do the libraries in the Metropole Européenne de Lille. As part of the LivreEtLecture project, an objective is to study the spatial integration of libraries into other types of public places and to evaluate the consequences on the use of libraries. For this purpose, we counted the number of each type of infrastructures around the libraries to establish 'spatial signatures' which we compared to each other according to the rate of active borrowers of the institutions. Then, interviews are conducted with the library users from different towns to refine the choice of the data and the scale of our studies. In this article, we describe the experiments into the libraries of Lille.

*MOTS-CLÉS :* Analyse spatiale, Bibliothèques, Système d'information Géographique.

*KEYWORDS:* Spatial analysis, Libraries, Geographical Information Systems.

---

SAGEO'2018 - Montpellier, 6-9 novembre 2018



## 1. Introduction

La bibliothèque est étymologiquement un lieu de conservation et de consultation des textes. C'est aussi plus généralement une institution publique dont « l'accès et les services sont gratuits et ont pour but ultime de répondre aux besoins d'information, de loisir, de culture et d'éducation de tous les individus d'une collectivité sociale, celle-ci étant définie en lien avec les résidents (citoyens) d'une municipalité » (Mittermeyer, 2004). C'est donc un lieu public, ouvert et accessible à tous à vocation culturelle mais aussi sociale. La définition originelle est restrictive et on observe aujourd'hui une volonté de changer l'image austère que l'on a de ce lieu de savoir et de promouvoir les bibliothèques comme lieu de sociabilité. Certains spécialistes parlent même aujourd'hui de la « bibliothèque troisième lieu ». Ce nouveau type de bibliothèque introduit des nouveaux usages sociaux (Servet, 2010). Les bibliothèques offrent alors la possibilité à leurs usagers d'y pratiquer des activités variées ; on n'y vient plus seulement pour consulter ou emprunter des ouvrages, ce sont de véritables lieux de vie qui font partie intégrante de la vie des citoyens.

Cependant malgré cette évolution des pratiques et usages des bibliothèques il existe depuis plusieurs années une baisse de leur fréquentation qui se traduit par une baisse du nombre d'inscrits et une diminution du nombre d'entrées physiques. Cela s'observe notamment pour les établissements centraux des communes de plus de 50 000 habitants. A l'inverse, les bibliothèques des communes de moins de 20 000 habitants semblent mieux résister à ce phénomène (Poissenot, 2010). La bibliothèque est souvent le seul équipement culturel permanent dans une commune, qui touche les habitants de toutes les tranches d'âge. C'est « le premier outil de démocratisation culturelle », mais il n'est pas facile pour une commune de faire vivre une bibliothèque, notamment à cause de l'investissement financier que cela représente (Pinard, 2009). Il est donc important que la fréquentation soit au rendez-vous et il est alors intéressant d'étudier les facteurs, y compris la situation géographique au sein du quartier voire de la commune, pouvant influencer la venue en bibliothèque ; l'approche géographique prend alors tout son sens.

La baisse de la fréquentation a également été observée dans la Métropole Européenne de Lille et c'est dans ce contexte qu'a été mis en place le projet LivreEtLecture qui s'intéresse notamment à la place de la bibliothèque dans la vie des citoyens<sup>1</sup>. Il est fait comme hypothèse de départ que la présence de certains types de lieux à proximité d'une bibliothèque favorisera les déplacements des citoyens vers elle. Le travail présenté ici détaille une méthodologie combinant une approche permettant (1) de mettre en avant l'insertion du réseau des bibliothèques sur le territoire sur la MEL en analysant leur positionnement par rapport à d'autres types de lieux (culturels, économiques, de loisirs et de

---

1. BiblioMEL, 2017, <https://bibliomel.hypotheses.org/proposition-scientifique>

services), (2) de croiser cela avec des données quantitatives de fréquentation et des données qualitatives résultantes d'entretiens pour tenter d'analyser les pratiques et les trajectoires des usagers. La liste des lieux est ensuite modifiée à partir des premiers résultats d'enquêtes auprès des usagers. Avant tout traitement des données, la remise en contexte du projet ainsi que les méthodes d'analyse spatiale et de recueil de données sont présentées en section 2. La section 3 permet de décrire la zone d'étude, les différentes données utilisées et la méthodologie mise en place et appliquée à la zone d'étude. Nous nous concentrons dans cet article sur les données correspondant au réseau des bibliothèques de la commune de Lille. La section 4 résume les expérimentations croisant des approches cartographiques et qualitatives afin d'évaluer la pertinence de cette méthodologie, ainsi que les perspectives à ces travaux, à commencer par étendre l'étude sur l'ensemble du réseau des bibliothèques de la MEL.

## 2. Contexte et enjeux de l'étude

### 2.1. *L'apport de l'approche géographique et cartographique*

Les bibliothèques sont dans certains territoires, notamment en province, les seuls services publics d'accès gratuits. Elles ont un rôle social et culturel important. De ce fait, leur situation géographique joue sur la qualité de vie des habitants d'un quartier. D'après le sociologue Edmond Preteceille, « Les bibliothèques font partie du réseau des équipements et services dont l'inégale distribution spatiale contribue à définir les qualités différentielles des espaces urbains et à déterminer ainsi les conséquences de la ségrégation sociale sur les modes de vie. » (Bertrand, 2001).

De nombreux facteurs peuvent influencer la venue en bibliothèque des usagers. Cécile Gobbo a étudié plus précisément les petites bibliothèques publiques en milieu urbain et met en avant plusieurs facteurs jouant sur leur fréquentation. L'une des raisons principales d'une plus forte fréquentation est la proximité au lieu d'habitation. L'auteur met également en avant le rôle que jouent les horaires et la signalétique extérieure sur la sélection du public (Gobbo, 2010). Des horaires d'ouverture limités restreignent le type de public. Une mauvaise signalétique ne favorisera quant à elle que la venue d'usagers connaissant déjà les lieux. L'auteur indique également que les bibliothèques situées dans des équipements municipaux ont plus de chance d'être connues du public. Le rapport de la directrice de la médiathèque départementale de la Drôme va dans ce sens et préconise plusieurs éléments à prendre en compte lors de l'implantation d'une bibliothèque pour assurer son attractivité : rechercher la proximité avec d'autres équipements ou services, prendre en compte la desserte par les différents moyens de transport, favoriser les modes de déplacements doux et l'usage des transports en commun, aménager les abords de l'équipement, veiller à une bonne visibilité du bâtiment et une bonne signalétique (Pinard, 2009).

Selon un rapport de la MEL analysant les résultats de plusieurs enquêtes, les usagers choisissent une bibliothèque plutôt qu'une autre car cela correspond aux moments les plus pratiques pour exercer une autre activité et car elle se situe à proximité d'autres équipements ou services utilisés par l'utilisateur (Lille, 2016). Les usagers couplent donc leur venue en bibliothèque avec d'autres activités de loisirs, culturelles ou sportives. La bibliothèque fait partie intégrante des dynamiques sociales d'une ville et en ce sens soulève les enjeux propres aux espaces publics, c'est-à-dire des enjeux de mobilité, des enjeux relatifs aux différents usages (civils, festifs, culturels, etc.), des enjeux de sociabilité, et un enjeu d'identité (Bassand, Güller, 2001).

Dans ce sens, nous faisons l'hypothèse que l'analyse de l'environnement spatial d'une bibliothèque peut nous aider à comprendre les facteurs qui jouent sur la fréquentation et l'attractivité de l'établissement.

## ***2.2. L'analyse spatiale, un outil pour étudier les localisations et les interactions entre objets***

L'espace géographique ne se construit pas au hasard, c'est un produit social (Capel, 1975) ; ce sont les déplacements et les différents types de flux (humains, de marchandises, d'information...) entre des lieux différents qui le construisent (Pumain *et al.*, 2001). L'intensité des interactions entre les lieux diffère selon le motif de déplacement et le mode de transport. Il existe des liens étroits entre certains types de flux et certains types de distance.

Dans le cadre du projet LivreEtLecture, on cherche à identifier des types d'organisations spatiales autour des bibliothèques qui pourraient expliquer leur « attractivité », mesurée ici grâce aux données quantitatives fournies par l'observatoire de la lecture publique et aux données des bibliothèques du réseau de Lille. D'après le géographe Hägerstrand, l'ensemble des lieux visités par une personne forme un espace d'interaction ; il correspond à un espace de mobilité dans lequel l'individu se déplace (Pumain, Saint-Julien, 2008). L'espace d'interaction dépend donc des lieux de résidence, de travail et des différents loisirs pratiqués par la personne. Nous nous intéresserons ici à l'espace de mobilité autour de la bibliothèque et aux activités pratiquées avant et après la venue en bibliothèque. L'étude menée ici ne s'intéresse pas exclusivement aux déplacements entre le domicile de l'utilisateur et la bibliothèque mais entre la bibliothèque et l'ensemble des autres lieux visités. Notre analyse est simplifiée car le système est en réalité plus complexe puisqu'il faudrait idéalement tenir compte des horaires d'ouverture des différents lieux. Ceux-ci ont un rôle important sur l'organisation de la journée des citoyens et leurs motifs de déplacement.

Ne pouvant nous pencher sur des mobilités individuelles, nous nous appuyons au cours de cette étude sur des tendances de déplacements, éventuellement associées à des groupes d'individus. L'échelle d'étude et le choix des lieux cartographiés seront détaillés section 3.

### 2.3. *Les cartes mentales, un outil pour recueillir les représentations spatiales des citoyens*

Les analyses spatiales permettent de mettre en évidence quels sont les lieux qui se ressemblent selon des critères d'organisation spatiale. Ces critères d'analyse sont simplifiés (pour des questions pragmatiques de calcul) et certains critères sont omis par choix subjectif, par oubli ou encore par difficulté d'intégration dans l'analyse. En ce sens, on peut se demander dans quelle mesure les analyses spatiales suffisent pour expliquer les comportements spatiaux des citoyens et plus particulièrement leur venue en bibliothèque. C'est pourquoi nous avons également mené des entretiens auprès des usagers et leur avons demandé de dessiner une carte mentale du quartier de la bibliothèque.

« La carte mentale est une représentation subjective de l'espace urbain par un habitant à partir des lieux qu'il a l'habitude de fréquenter » (Merlin, Choay, 1988). Cette définition souligne la subjectivité de la représentation spatiale que les individus ont de l'espace qui les entoure en fonction de leur expérience. D'après Horacio Capel, les individus possèdent des « schémas mentaux de l'espace, dans lesquels les éléments distincts sont localisés en fonction de leur position relative. » (Capel, 1975). Ces schémas mentaux peuvent être à la fois individuels et collectifs (Fournand, 2003). Ainsi, de nombreux géographes, psychologues et sociologues se sont penchés sur le dessin de cartes mentales, pour obtenir une représentation graphique de l'espace vécu par les citoyens. L'analyse de ces représentations permettrait d'expliquer les comportements des individus (Breux *et al.*, 2010).

Les cartes mentales dessinées, autrement appelées « sketch map » révèlent donc une perception individuelle et subjective de l'espace (Capel, 1975). Elles permettent une spontanéité de la réponse, une grande liberté de la personne interrogée en passant outre la barrière de la langue et du langage en général et peuvent être utilisées auprès d'individus appartenant à toutes les classes d'âge. Cependant de nombreux auteurs soulèvent la difficulté d'interprétation des dessins due à leur hétérogénéité ainsi que les biais causés par les qualités de dessin de chacun qui peuvent être un obstacle à la représentation des schémas mentaux (Breux *et al.*, 2010). Cet outil par la singularité des informations qu'il permet de recueillir est utilisé dans de nombreux domaines ayant des objectifs variés. La méthodologie employée par les différents chercheurs reste relativement identique d'une étude à l'autre. Elle consiste à demander à un individu de dessiner sur une feuille blanche avec un crayon noir un espace donné. Les sketch map sont souvent accompagnées d'un questionnaire ou d'un entretien qui permettent d'obtenir « une information supplémentaire sur les valeurs symboliques et/ou sentimentales » de l'espace (Rowntree, 1997).

Dans le cadre de notre étude, l'intérêt de demander aux usagers de dessiner une carte mentale est de pouvoir déterminer la connaissance de l'environnement (à l'échelle du quartier) de la bibliothèque à partir de leur expérience et de

leurs activités dans ce même environnement. Cela permettra de faire ressortir les éléments constitutifs du paysage qui sont communs aux différents usagers de la bibliothèque quel que soit leur âge ou leur origine sociale (Lynch, 1960). Nous tirerons de ces dessins les lieux les plus souvent fréquentés ainsi que les limites de l'espace parcouru. On cherche aussi à identifier quelle place les usagers accordent à la bibliothèque dans leur quartier.

### 3. Matériel et méthodes

Nous émettons donc l'hypothèse que l'environnement spatial, et plus précisément la présence de certains types de lieux dans le voisinage de la bibliothèque, a une influence sur la fréquentation de la bibliothèque. Nous partons de l'analyse spatiale pour dégager des corrélations entre voisinage des établissements et taux de fréquentation similaires. Pour cela, nous définissons pour chaque bibliothèque sa « signature spatiale », i.e. l'ensemble des lieux d'intérêts localisés dans une zone de chalandise, et nous corrélons ces signatures aux taux de fréquentation pour tenter d'expliciter des liens entre le bon fonctionnement d'une bibliothèque et les différents types de lieux qui l'entourent. Puis nous croisons ces résultats avec des enquêtes auprès des usagers qui nous permettront d'affiner le choix des analyses et de faciliter l'interprétation des résultats cartographiques.

#### 3.1. La MEL et son réseau de bibliothèques

La Métropole Européenne de Lille est constituée de 90 communes. Elle est la deuxième agglomération française par sa densité de population et la quatrième par sa taille. La MEL est polycentrique, les communes de Lille, Tourcoing, Roubaix et Villeneuve d'Ascq rassemblent à elles seules 40% des métropolitains en comptant chacune plus de 65 000 habitants (Lille Métropole). D'après les résultats d'enquêtes de la Métropole de Lille, les ouvriers et agriculteurs sont sous-représentés parmi les usagers des bibliothèques. A l'inverse, les employés, professions intermédiaires, les cadres et professions intellectuelles supérieures sont sur-représentés. On observe également que la fréquentation de la bibliothèque varie tout au long de la vie des individus. En effet, les usagers de moins de 15 ans et de plus de 61 ans sont sous-représentés tandis que les 30-44 ans sont les plus représentés parmi les usagers. Ainsi la composition de la population joue sur la fréquentation des bibliothèques. La MEL se caractérise par une population plus jeune, un revenu médian plus faible et un taux de chômage plus élevé que la moyenne nationale. Il existe également de fortes inégalités territoriales au sein de la MEL. Certaines communes très pauvres telles que Roubaix (45% de sa population vit en dessous du seuil de pauvreté) jouxtent d'autres communes telles que Croix, Marcq-en-Baroeul et Bondues où le nombre de contribuables à l'ISF est très élevé. L'agglomération de Lille est la deuxième agglomération française de plus de 250 000 habitants la plus ségréguée après la

région parisienne et « 6 des 10 quartiers de la ville de Lille font partie des « quartiers prioritaires » des politiques de la ville » (Degeyter, 2017). Cette ségrégation est en partie due à l'histoire de la ville et à la désindustrialisation qu'elle a subie à partir des années 1960. Un autre public potentiel des bibliothèques à ne pas négliger est constitué des 105 000 étudiants de toutes formations supérieures confondues qui représentent plus de 10% de la population en 2015 (Lille Métropole).

Dans cet article, nous nous concentrons sur le réseau des 9 bibliothèques de quartier de Lille, avec comme tête du réseau la bibliothèque Jean Lévy située au centre-ville et une bibliothèque par quartier (cf. figure 1). La bibliothèque de Lomme (commune voisine de Lille) a quant à elle rejoint le réseau en juillet 2016. Les différentes caractéristiques de ces bibliothèques sont détaillées dans le tableau 1<sup>2</sup>.

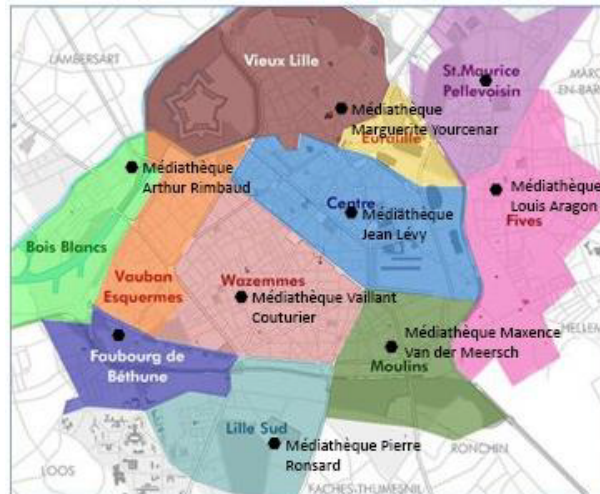


FIGURE 1. Plan des quartiers de Lille et localisation des bibliothèques municipales du réseau.

On observe une certaine hétérogénéité des infrastructures dans les différents quartiers. Certains bénéficient de bibliothèques plus grandes que d'autres (tout en ayant le même nombre d'habitants) mais cela n'a pas l'air d'influencer le nombre d'inscrits dans chacun des établissements. Ainsi, les bibliothèques de Wazemmes ou de Fives ayant une plus petite surface accueillent autant d'inscrits que la bibliothèque de Moulines qui est deux fois plus grande.

2. données quantitatives fournies par l'Observatoire de la lecture publique qui dépend du ministère de la Culture et de la Communication

TABLE 1. *Caractéristiques des bibliothèques du réseau de Lille, chiffres de mai 2017*

Bibliothèque	Surface (m <sup>2</sup> )	Nb inscrits	Nb entrées	Documents	Population quartier
Bois Blanc	546	1 147	2 526	27 372	7 160
Faubourg de Béthune	695	1 172	3 371	33 535	7 760
Fives	413	1 866	4 571	24 977	19 060
Jean Lévy	1 300	7 429	16 245	457 589	23 000
Lille Sud	1 086	1 701	4 401	28 901	19 300
Moulins	888	1 964	NC	43 135	19 800
Saint Maurice Pellevoisin	552	2 656	6 438	22 956	16 690
Vieux Lille	713	1 877	NC	37 816	18 840
Wazemmes	327	1 846	5 037	24 909	26 610

### 3.2. Données utilisées

Pour cette étude nous nous appuyons sur plusieurs types de données issues de fournisseurs nombreux et variés (Open Data MEL, service SIG MEL, Open Street Map, réseau Transpole, et base de données des chercheurs du projet) : des données quantitatives liées aux usagers des bibliothèques, et des données de localisation des bibliothèques et de différents lieux d'intérêt. On appelle lieux d'intérêt, les lieux où les usagers sont susceptibles de se rendre avant ou après leur venue en bibliothèque. En collaboration avec l'équipe du service Culture de la MEL, les lieux d'intérêt sélectionnés pour l'étude sont les points d'accès aux moyens de transports (arrêts de bus, stations de métro, gares, zones de stationnement), les lieux culturels sous-divisés en deux ensembles (les lieux de pratique quotidienne : maisons de quartier, écoles de musique, librairies, etc. et les lieux de sortie culturelle fréquentés de façon plus occasionnelle : musées, cinémas, théâtre, etc.). Les premiers entretiens réalisés auprès des usagers ont révélé que ceux-ci profitaient de leur passage à la poste ou à la mairie pour passer à la bibliothèque et y rendre des documents empruntés par exemple. Suite à cela, il a été décidé d'ajouter les bâtiments administratifs (mairies, palais de justice, bureaux de poste) à la liste des lieux d'intérêt.

Au début du projet, il n'existait pas de base de données spatialisée complète regroupant l'ensemble des lieux culturels et autres lieux d'intérêt de la métropole de Lille. La première étape de l'étude a donc été de récolter et de géocoder les données pertinentes à partir de sources hétérogènes (cf. Figure 2). Pour compléter les données fournies par la MEL (liste des écoles, collèges, lycées, établissements d'enseignement supérieur, liste des lieux culturels de type cinéma, théâtre, fabrique culturelle, salle d'exposition, salle de spectacle, liste des parcs et jardins), nous nous sommes notamment appuyé sur OpenStreetMap qui, bien que non exhaustive et non fiable à 100%, nous a permis d'identifier des lieux 'alternatifs', comme par exemple les petites salles de spectacle ou salle de projection associatives. Nous avons également pu récupérer une bonne partie des commerces de proximité (boulangeries, épiceries, marchés, super-

marchés). L'avantage d'OSM est que la communauté étant active, les données sont actualisées régulièrement et on peut y trouver des commerces ayant vu le jour il y a à peine quelques mois par exemple. Certains lieux tels que les points de vente du livre, les maisons de quartier et MJC (Maison des Jeunes et de la Culture) ont été fournis par des membres du projet, menant des études de terrains dans ces lieux.

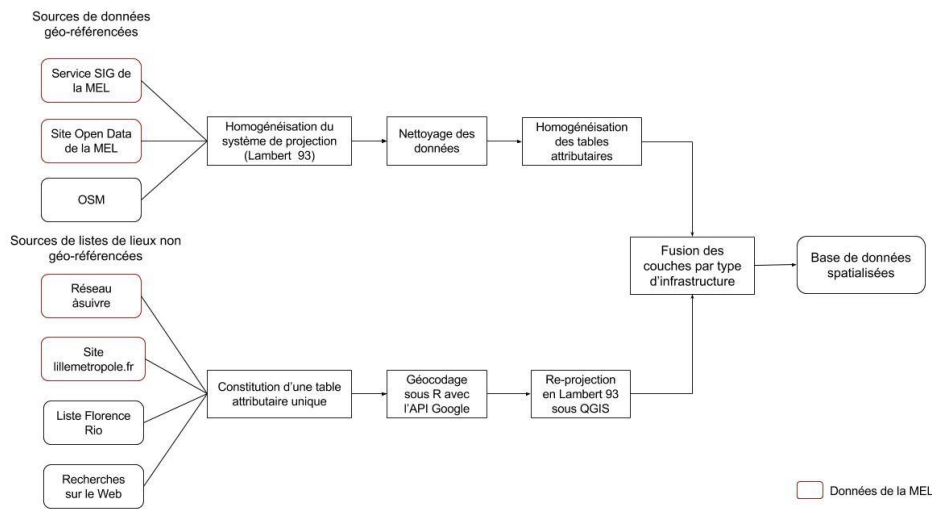


FIGURE 2. *Prétraitement des données géographiques.*

### 3.3. Caractérisation de la zone de chalandise des bibliothèques

La zone de chalandise d'une bibliothèque est la zone géographique où sont implantés ses publics potentiels. Il existe de fortes variations de ce périmètre selon la démographie, la composition sociale, les modes de transport, etc. (Asselin, 2016). La zone de chalandise d'une activité peut se définir selon différents critères :

- selon les informations recueillies auprès des clients (adresse fournie lors de l'abonnement par exemple),
- selon le temps de parcours : isochrone. « une zone isochrone est une surface délimitée par une courbe géométrique délimitant les points accessibles par un véhicule – terrestre ou aérien – en un temps donné »,
- selon la distance à vol d'oiseau : distance euclidienne<sup>3</sup>.

3. Owlapps, 2012, <http://www.owlapps.net/articles/zones-de-chalandise-zones-isochrones-et-zones-disodistance>



Ne disposant pas ici des informations sur les usagers des bibliothèques, nous nous orientons sur les deux dernières méthodes. Concernant les distances basées sur le temps de parcours ou sur la distance à vol d'oiseau, l'enquête sur les bibliothèques municipales de Paris menée en 2003 précise que « la fréquentation et la connaissance des bibliothèques diminuent sensiblement au-delà de 800 m de distance » et « à partir de 800m séparant une bibliothèque du domicile, le nombre d'inscrits diminue passant de 33% à 25% » (Yves, 2005). Cela nous permet donc d'obtenir une première distance objective d'étude. Au cours des premiers entretiens réalisés auprès des usagers des bibliothèques il leur est demandé leur temps de parcours pour venir à la bibliothèque. Pour beaucoup, ce temps est inférieur à 10 minutes. C'est pourquoi, dans le cas de l'étude du réseau des bibliothèques de Lille, situé en milieu exclusivement urbain dense où la plupart des déplacements se font à pied et sur de plus courtes distances, nous testons les analyses à 800m mais aussi à deux échelles plus fines, à savoir 200m et 400m.

### **3.4. Une méthodologie en deux étapes**

#### *3.4.1. Analyse spatiale du réseau des bibliothèques*

L'objectif est ici de créer un indicateur qui permette de caractériser l'environnement spatial afin de tester s'il existe une corrélation entre l'aménagement autour des bibliothèques et leur fréquentation. Cet indicateur s'appuie sur le type d'infrastructures dans le voisinage de la bibliothèque et la diversité des infrastructures. Nous utilisons un script qui permet compter le nombre de chaque type d'infrastructures dans un certain rayon autour de la bibliothèque. Les résultats sont ensuite assemblés et exportés sous forme de tableau dans lequel à chaque bibliothèque est assigné le nombre d'infrastructures par type, i.e. sa « signature spatiale ». Nous pouvons ainsi facilement obtenir la diversité d'infrastructures présentes dans la zone tampon, en dénombrant les catégories pour lesquelles le nombre de lieux n'est pas nul.

#### *3.4.2. Les enquêtes auprès des usagers des bibliothèques*

L'objectif est ici d'identifier l'ensemble des lieux fréquentés par les usagers des bibliothèques et de comparer cette liste à celle établie précédemment avec les membres du projet. D'après Colette Cauvin, il existe différentes méthodes permettant d'identifier des lieux. Ainsi, une enquête « libre, spontanée » fait appel à la mémoire et à l'imagination simultanément, tandis que la proposition d'une liste de lieux préétablie que l'enquêté doit cocher selon ses habitudes ne fait appel qu'à la mémorisation. Chacune des méthodes présente des biais ; la première peut entraîner des oublis de la part de l'enquêté tandis que la deuxième s'appuie sur un choix arbitraire de l'enquêteur et risque de restreindre l'enquêté dans ses choix et de passer à côté de lieux auxquels l'enquêteur n'aurait pas pensé. C'est pourquoi ces deux méthodes sont parfois utilisées de façon complémentaire (Cauvin, 1999). C'est ce que nous avons essayé de reproduire en

demandant dans un premier temps aux usagers de dessiner leur quartier / lieu de vie en situant la bibliothèque. Le fait de demander à l'utilisateur de dessiner une échelle et de lui demander de placer la bibliothèque sur son dessin permet d'avoir un point commun entre le système spatial géo-référencé et les représentations cognitives individuelles. Cet exercice cartographique est intégré dans un entretien qui permet de connaître le profil de l'utilisateur et ses habitudes. Les objectifs de cet exercice graphique sont notamment de déterminer quels sont les lieux connexes à la bibliothèque les plus fréquentés, ceux qui sont évités, d'identifier les points de repères autour de la bibliothèque et de dégager les limites de la zone d'attractivité de la bibliothèque (i.e. les zones "sombres" non dessinées ou très schématiquement et correspondant aux limites de l'espace parcouru par l'enquête). On ne cherche pas ici la représentativité, ainsi nous veillons à interroger un échantillon varié d'utilisateurs de la bibliothèque représentant plusieurs tranches d'âge. Nous essayons aussi d'aller à la bibliothèque deux jours différents de la semaine car les publics ne sont pas toujours les mêmes. Par exemple, le mercredi après-midi les utilisateurs surreprésentés sont les mères de famille avec enfants, tandis qu'un autre jour en matinée, on observera plutôt la présence de personnes âgées.

Les enquêtes auprès des utilisateurs des bibliothèques nous permettent entre autres d'évaluer la pertinence des lieux choisis pour la constitution des indicateurs et l'échelle choisie.

#### 4. Expérimentations

##### *4.1. Croisement des analyses cartographiques et des données quantitatives de fréquentation*

Nous comparons ici les signatures spatiales aux différentes échelles de 200m, 400m et 800m pour les bibliothèques du réseau de Lille. La figure 3 présente ces signatures spatiales pour chaque bibliothèque du réseau, classées selon le nombre d'entrées de mai 2017 (par ordre décroissant du haut vers le bas). Les signatures spatiales sont comparées visuellement et nous observons qu'elles sont similaires entre elles pour des bibliothèques en milieu urbain dense (Lille) quelle que soit l'échelle. Seule la médiathèque de Lille Sud présente une signature spatiale différente entre 400m et 800m. On observe également qu'à 800m les voisinages chaque médiathèque semblent présenter la même organisation spatiale, en dehors des bibliothèques Jean Lévy et Vaillant Couturier (Wazemmes) proches desquelles se trouvent des lieux de pratique culturelle. Ces analyses ne nous permettent pas d'identifier clairement une signature spatiale générique caractérisant un ensemble de bibliothèques qui fonctionnent. Nous menons actuellement des analyses similaires sur l'ensemble du réseau des bibliothèques de la MEL pour tenter d'identifier une ou plusieurs signature(s) spatiale(s) d'intérêt. Des entretiens auprès des utilisateurs sont alors menés afin de déterminer si les lieux identifiés et les échelles d'étude sont pertinents.

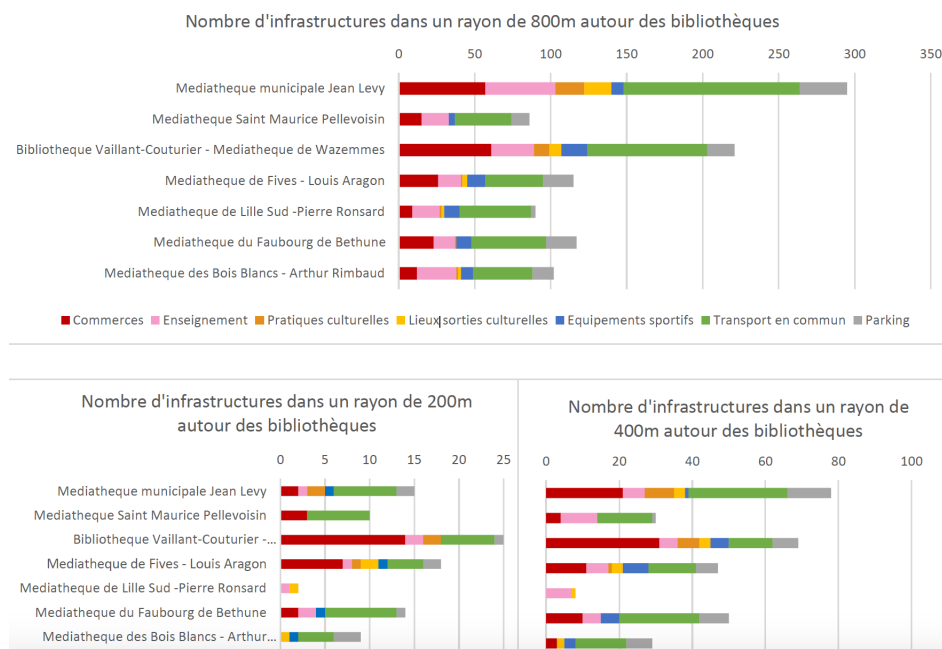


FIGURE 3. Prétraitement des données géographiques.

#### 4.2. Résultats d'enquêtes dans les bibliothèques municipales de Lille

Nous décrivons ici les résultats d'enquêtes menées au sein de deux bibliothèques du réseau en mai/juin 2017, à savoir 10 entretiens menés à la bibliothèque centrale du réseau nommée Jean Levy et 15 menés à la bibliothèque Vaillant Couturier localisée dans le quartier populaire Wazemmes, dans laquelle on retrouve toutes les classes d'âge et toutes les catégories socio-professionnelles. Ce nombre d'entretiens trop faible ne permet pas encore de dégager de tendances mais permet de tester la grille d'entretien et la réceptivité des enquêtés aux outils que sont les cartes mentales. Le projet étant toujours en cours, la phase d'entretiens sera poursuivie et étendue à l'ensemble du réseau dans les mois à venir.

Concernant les entretiens menés à Jean levy, les personnes interrogées viennent aussi bien du centre-ville (moins de 5 minutes à pied) que d'autres quartiers (Wazemmes, Vieux Lille). Cette bibliothèque est appréciée pour sa taille, le nombre de documents qu'elle offre, l'accueil du personnel ainsi que pour son ambiance calme. La présence du petit jardin à l'arrière du bâtiment est également très appréciée, à tel point qu'une personne n'a représenté que cela sur sa carte mentale, revenant spécialement le jour de l'entretien pour en profiter

par beau temps. Les usagers interrogés ont du mal à positionner les bâtiments et rues du quartier. Il semble manquer un point de repère comme c'est le cas pour la bibliothèque de Wazemmes (place du marché de Wazemmes).

Concernant la bibliothèque municipale de Wazemmes, les enquêtés appréhendent la situation géographique de la bibliothèque, la facilité de stationnement à proximité, ses collections et son calme. Cependant plusieurs habitués déplorent l'insécurité du quartier ainsi que la présence de dealers et critiquent l'insuffisance de l'intérêt porté par les politiques publiques sur le quartier. Certains usagers ont même dessiné les dealers sur leur carte mentale. Deux cartes mentales d'enquêtés de la bibliothèque de Wazemmes sont présentées figure 4. Au cours des entretiens, il s'est avéré que très peu d'usagers corrélaient leur venue en bibliothèque avec une autre activité ou alors n'en ont pas conscience. Souvent les usagers ne viennent que pour la bibliothèque qui constitue une activité "riche en soi". Leur venue est planifiée et ils se rendent à la bibliothèque à pied. Beaucoup utilisent le catalogue en ligne pour réserver des documents d'autres bibliothèques du réseau et les faire venir grâce au service de navettes proposé. Ce service leur permet de bénéficier de l'ensemble du fond du réseau des bibliothèques de Lille sans avoir à se déplacer d'une bibliothèque à l'autre.

FIGURE 4. *Sketch Map de deux enquêtés de la bibliothèque de Wazemmes.*

### 4.3. Discussion

La généralisation des résultats et des entretiens s'avère encore très difficile du fait du peu d'entretiens menés et de l'hétérogénéité des profils d'utilisateurs. Tous n'ont pas le même rapport à la bibliothèque et n'y cherchent pas la même chose : objectifs de la venue, durée du séjour, etc. Pour cette enquête, l'utilisation de l'outil des cartes mentales ne semble pas le plus efficace pour déterminer les lieux les plus fréquentés autour de la bibliothèque, mais permet d'enrichir les réponses aux entretiens et ainsi de compléter l'analyse qualitative. En effet, les lieux les plus souvent cités comme étant visités avant ou après la venue en bibliothèque sont des lieux fréquentés quasi-hebdomadairement, tels que les commerces par exemple. Les enquêtés ne pensent pas à ce type de lieu quand ils se retrouvent devant la feuille blanche et privilégient les points de repères marquants (église, place, parc, etc.) dans leur dessin. Les questions posées lors de l'entretien semblent plus efficaces pour obtenir ce type d'information. On pourrait également envisager de présenter un plan du quartier aux personnes interrogées et leur demander d'entourer les lieux qu'ils fréquentent le plus avant ou après être venu en bibliothèque. D'après l'enquête menée, les personnes interrogées ne semblent pas privilégier une bibliothèque par rapport à une autre selon son voisinage. L'échelle d'étude est donc difficile à fixer. Les quelques comparaisons qui ont été tirées par les enquêtés montrent qu'ils privilégient une bibliothèque à une autre selon sa taille, son fond documentaire, son accueil ou son ambiance. Plusieurs usagers déclarent beaucoup aimer les bibliothèques de Marcq-en-Baroeul et de Lomme (qui fait désormais partie du réseau). Ces bibliothèques sont décrites comme modernes, agréables et « donnant envie de rester ». Aucun n'a évoqué les infrastructures situées dans le quartier de la bibliothèque comme atout, comme nous l'avions présumé dans notre hypothèse de départ suite aux différentes lectures effectuées. Ainsi, l'intérieur de la bibliothèque semble avoir plus d'incidence que l'organisation extérieure du quartier. Cependant comme nous avons pu le constater avec les usagers de Wazemmes, l'insécurité d'un quartier joue tout de même un rôle important et peut dissuader les usagers de venir. Certains enquêtés ont également évoqué leur réticence à aller dans certaines bibliothèques à cause des mauvais rapports avec un membre du personnel. Les usagers viennent dans les bibliothèques de quartier car ce sont les plus proches de leur lieu de résidence mais apprécient également de pouvoir se rendre à la bibliothèque Jean Lévy qui est plus spacieuse et propose un plus grand choix de documents. Quelque soit l'environnement autour de la bibliothèque, ce n'est pas un certain type d'infrastructures ou l'architecture du bâtiment qui vont inciter les usagers à s'y rendre, mais bien ce qui est proposé à l'intérieur. Ces résultats ont été obtenus suite aux enquêtes menées dans les bibliothèques de Wazemmes et du centre-ville de Lille. Ils restent néanmoins à approfondir en ciblant plus précisément chacune des catégories d'utilisateurs, travail que nous avons initié à plus grande échelle depuis sur les neuf bibliothèques de la MEL.

Une fois la méthodologie stabilisée, il serait intéressant de travailler à l'extension de l'étude sur l'ensemble du réseau de la MEL dont la répartition inégale semble suivre celle la densité de population le long de l'axe Lille-Tourcoing.

## 5. Conclusion

Le travail fourni a permis dans un premier temps de constituer une base de données spatiale de différents lieux de la MEL. Il a également permis de tester l'utilisation d'outils tels que les cartes mentales et entretiens, pour compléter des analyses spatiales.

L'étude menée apporte ainsi quelques éléments de réflexion sur la pertinence de la méthodologie de l'analyse cartographique et plus particulièrement sur le choix des données et de l'échelle d'étude. L'approche cartographique a pour objectif d'expliquer l'attractivité qu'exerce une bibliothèque par rapport à une autre à partir de l'analyse des interactions entre les différents lieux fréquentés par les citoyens de la ville de Lille. Les premiers résultats, confrontés à des entretiens menés auprès des usagers de différentes bibliothèques municipales, ont permis de mettre en évidence une réalité autre que celle qui avait été présumée. Plusieurs bibliothèques présentant une organisation de quartier similaire peuvent avoir des taux de fréquentation différents, et inversement, des bibliothèques situées dans des quartiers aux organisations spatiales différentes peuvent fonctionner en terme de fréquentation aussi bien l'une que l'autre.

Nous sommes partis d'hypothèses pour mettre en place une méthodologie d'analyse spatiale que nous avons ensuite comparée aux réponses d'entretiens des usagers. Au vu des premiers résultats, l'hypothèse de départ mériterait d'être affinée afin de considérer d'autres facteurs liés aux déplacements des usagers tels que les horaires d'ouverture des différents lieux ou les caractéristiques sociologiques des usagers. Il serait aussi intéressant de dégager des comportements spatiaux par type d'usagers plutôt que d'essayer de tirer une généralisation. Creuser les liens entre les différents acteurs en prenant en compte dans l'étude l'influence des différents partenariats mis en place entre bibliothèques et autres structures (associations, maisons de quartiers, lieux culturels, écoles, etc.) sur les usagers et leur venue en bibliothèque apporterait également un éclairage nouveau sur les pratiques des usagers.

## Bibliographie

- Asselin E. (2016). Evaluer l'impact de sa bibliothèque sur le territoire : mesurer la zone de chalandise ou rayonnement de sa bibliothèque. In *Évaluer la bibliothèque par les mesures d'impacts*. Presses de l'Enssib. Villeurbanne. La boîte à outils, p. 37.
- Bassand M., Güller P. (2001). *La dynamique des espaces publics*. Lausanne, Vivre et créer l'espace public, Presses polytechniques et universitaires romandes.

- Bertrand A.-M. (2001). *Les bibliothèques municipales et leurs publics: pratiques ordinaires de la culture*. Paris, Bibliothèque publique d'information, Centre Pompidou. Etudes et recherche.
- Breux S., Reuchamps M., Loiseau H. (2010). *Apports et potentialités de l'utilisation de la carte mentale en science politique*. Transeo Review [en ligne]. Figurer l'espace en sciences sociales.
- Capel H. (1975). L'image de la ville et le comportement spatial des citadins. In *Espace géographique*, vols. 4, n° 1, p. 73-80.
- Cauvin C. (1999). *Pour une approche de la cognition spatiale intra-urbaine*. Cybergeog.
- Degeyter C. (2017). *Sociologie de Lille*. Paris, La découverte.
- Fournand A. (2003). Images d'une cité. Cartes mentales et représentations spatiales des adolescents de Garges-lès-Gonesse. In *Annales de Géographie*, vols. 112, n° 633, p. 537-550.
- Gobbo C. (2010). *Les petites bibliothèques publiques en milieu urbain*. Thèse de doctorat non publiée, Mémoire d'étude - Diplôme de conservateur des bibliothèques. Ensib.
- Lille M. E. de. (2016). *Etude de fréquentation et d'impact portant sur l'adaptation des horaires d'ouverture de plusieurs groupements de bibliothèques / médiathèques*. Lille.
- Lynch K. (1960). *L'image de la Cité*. Dunod.
- Merlin P., Choay F. (1988). *Dictionnaire de l'urbanisme et de l'aménagement*. Paris, PUF.
- Mittermeyer D. (2004). La bibliothèque publique comme lieu citoyen : variations sur un thème. In *Documentation et bibliothèques*, vols. 50, n° 4, p. 265-272.
- Pinard J. (2009). Créer une bibliothèque projet culturel et projet spatial. In *Lire en Drome [en ligne]*, vol. Hors-Série.
- Poissenot C. (2010). La fréquentation en questions. In *Bulletin des bibliothèques de France (BBF)s*, vol. N° 5, p. 67-72.
- Pumain D., Saint-Julien T. (2008). *L'analyse spatiale*. Paris, Localisations dans l'espace, A. Colin. Cursus.
- Pumain D., Saint-Julien T., Dumas E., Mathian H. (2001). *Les interactions spatiales: flux et changements dans l'espace géographique*. Paris, Armand Colin.
- Rowntree B. (1997). Les cartes mentales, outil géographique pour la connaissance urbaine. Le cas d'Angers (Maine-et-Loire). In *Annales de Géographie*, vols. 176, n° 1, p. 585-604. Norois.
- Servet M. (2010). Les bibliothèques troisième lieu. In *Bulletin des bibliothèques de France (BBF)s*, vol. N° 4, p. 57-63.
- Yves A. (2005). L'enquête sur le public des bibliothèques de la Ville de Paris. In *Publics : quelles attentes ? Bibliothèques : quelles concurrence ? Éditions de la Bibliothèque publique d'information*, p. 25 - 28.

---

# Appariement automatique de données hétérogènes: textes, traces GPS et ressources géographiques

Amine Medad<sup>1</sup>, Mauro Gaio<sup>1</sup>, Sébastien Mustière<sup>2</sup>

1. LIUPPA, Université de Pau et des Pays de l'Adour, 64013 Pau Cedex, France

*im.medad@univ-pau.fr,mauro.gαιο@univ-pau.fr*

2. UPE, IGN ENSG, LASTIG COGIT

*Sebastien.Mustiere@ensg.eu*

---

*RÉSUMÉ. Les travaux que nous présentons dans cet article sont réalisés dans le cadre du projet ANR Choucas. Nous proposons une approche pour l'appariement automatique de traces, de textes de description de randonnées et de ressources géographiques (gazetiers et bases de données géographiques). L'objectif de cet article est d'exposer les premiers éléments d'une méthodologie d'appariement automatique dont les trois étapes sont : l'annotation des traces GPS, l'identification des entités nommées spatiales dans des textes décrivant des itinéraires de randonnées, et la mise en correspondance des toponymes.*

*ABSTRACT. The work we present in this article is carried out within the Choucas ANR project. We propose an approach for the automatic matching of traces, hiking description texts and geographical resources (gazetteers et geographical data bases). The main goal of this article is to expose the first methodological elements for the automatic matching of hiking descriptions and GPS traces. To achieve this goal three steps are needed: GPS traces annotation, spatial named entities identification and toponyms matching.*

*MOTS-CLÉS : Extraction d'information, traitement automatique du langage, données réparties et hétérogènes, appariement*

*KEYWORDS: Information retrieval, Natural language processing, heterogeneous and distributed data, matching*

---



## 1. Introduction

Grâce aux plateformes de partage d'informations de randonnées (VisoRando, Camp2camp, IGNRando, ... etc.), nous avons de plus en plus accès pour chacune d'entre elles à une description textuelle de l'itinéraire à suivre souvent accompagnée d'une trace GPS. Les auteurs de ces textes utilisent dans leurs descriptions des points de repères faciles à identifier mais qui compte tenu de leurs caractéristiques en termes de taille, de notoriété ou alors de la façon de les nommer, peuvent ne pas être présents dans les ressources géographiques.

Nos travaux de thèse ont pour but de répondre à la problématique posée par le projet ANR Choucas et plus particulièrement la structuration de données issues de sources textuelles hétérogènes, qui est présentée et expliquée dans la section 2 de l'article (Olteanu-Raimond *et al.*, 2017). Pour ce faire nous cherchons à mettre en correspondance des données hétérogènes : textes narratifs, traces GPS et ressources géographiques (gazetiers, BD géographiques et LOD « données ouvertes liées »), la solution à cette problématique implique la nécessité de répondre à plusieurs verrous scientifiques d'ordre :

1. Cognitif et linguistique : la complexité de la description via le langage naturel de concepts spatiaux décrivant des situations spatiales statiques (ex : description de la position d'une victime) ou dynamiques (ex : description d'itinéraire par les randonneurs) est un problème qui nécessite d'établir une passerelle entre les domaines de la linguistique, des sciences cognitives, et de l'informatique (traitement automatique du langage naturel). La réponse à cette problématique nécessite la conception de motifs pour identifier les indices linguistiques de relations spatiales dans des textes de randonnées.

2. Raisonnement spatial : peu de travaux s'intéressent à l'appariement de données issues d'un texte (l'itinéraire décrit) avec des données spatiales (la trace GPS). La randonnée peut être vue comme une suite de relations spatiales entre points de repères impliqués dans des actions de mouvement, cela implique d'une part, une modélisation opératoire des actions de mouvement, des relations et des entités extraites du texte et un choix non triviale du modèle de définition des relations spatiales le plus adapté, et d'autre part, la mise en place des mécanismes d'analyse spatiale permettant de raisonner sur des informations souvent à caractère implicite ou incomplet.

Dans cet article nous décrivons les premières étapes d'une méthode permettant de faire l'appariement entre les informations contenues dans les textes et les informations issues des traces GPS associées. L'appariement permettra de faire correspondre à chaque point de repère évoqué dans le texte une géolocalisation ou un ensemble de géolocalisations obtenues à partir de la trace GPS associée. En d'autres termes pour chaque point de repère : soit la localisation la plus plausible obtenue à partir des ressources, soit une localisation approximative pour ceux qui ne sont initialement pas répertoriés dans les ressources géographiques. Ainsi, ce travail permettra à terme de fournir des données de

randonnées enrichies (trace annotée avec des éléments textuels de localisation issu des textes de randonnées) et d'enrichir des bases de données de points de repères en montagne.

L'article est organisé comme suit : la partie 2 présente les données (textes, traces GPS et ressources géographiques) sur lesquelles va porter notre travail. Dans la partie 3 nous expliquons comment nous comptons faire l'appariement entre ces différentes données. La partie 4 est dédiée à une expérimentation sur une première tentative d'appariement. Dans la partie 5 nous concluons et proposons des perspectives.

## 2. Description des données manipulées

### 2.1. Les textes

Les textes décrivant des randonnées sont de type narratif. Ils se présentent sous la forme d'un discours comportant des descriptions spatiales. Ces descriptions sont essentiellement composées d'indications sur le chemin à prendre pour atteindre les différentes étapes constitutives de l'itinéraire à parcourir. Les indications n'apparaissent pas forcément dans l'ordre dans lesquelles les étapes seront réalisées. L'exemple ci-après montre un extrait de texte décrivant une randonnée :

*" À partir du port de Carantec, traverser la voie submersible. L'île étant tout en longueur, il suffit de suivre le chemin principal jusqu'au bout de l'île puis de revenir sur ses pas. L'île abrite quelques maisons mais surtout une superbe petite chapelle visible de loin car perchée sur une butte ... "*

L'indication donnée dans le texte est fréquemment formulée sous la forme d'une phrase contenant un verbe de mouvement, une ou plusieurs relations spatiales et une ou plusieurs entités nommées spatiales étendues (ENSE)<sup>1</sup> (Gaio, Moncla, 2017), le sujet étant généralement le randonneur. Dans la section 3 nous présentons comment nous comptons extraire les ENSE à partir de ces textes.

### 2.2. Les traces GPS

Les plateformes de partage d'informations de randonneurs associent souvent une trace GPS à chaque description de randonnée. Les traces GPS peuvent être

---

1. ENSE est une entité construite à partir d'un nom (propre ou commun) et d'un ou plusieurs concepts relatifs à l'expression de la localisation dans la langue, plusieurs niveaux d'encapsulation ont été définis chaque niveau ( $n$ ) est encapsulé dans le niveau précédent ( $n - 1$ ), ex : Paris est une ENSE de niveau 0, Mairie de Paris est une ENSE de niveau 1. Dans notre contexte de randonnée, les points de repères, lieux et point de passages sont considérés comme des ENSE

interprétées comme une autre représentation des déplacements décrits dans les textes en langage naturel. Elles sont souvent disponibles sous la forme de fichiers au format GPX<sup>2</sup> contenant obligatoirement une collection de coordonnées utilisables sous la forme de points de cheminement, trace ou itinéraire et de manière optionnelle d'autres informations.

Les traces GPS disponibles sur les plateformes contiennent des coordonnées 2D (Latitude, Longitude), parfois 3D (Latitude, Longitude, Altitude), avec ou sans horodatage (timestamp), et en général aucune autre information. Il est également important de souligner que les traces sur lesquelles nous travaillons sont en général des traces corrigées, où les positions aberrantes (Buard *et al.*, 2015) ont été supprimées au préalable.

### 2.3. Les ressources géographiques

Les ressources géographiques que nous utilisons se présentent sous la forme de gazetiers (Geonames, OpenStreetMap, GoogleMaps), de base de données géographiques (BD Topo, Base adresse nationale), et de LOD (DBPedia). Les informations dans ces ressources sont parfois redondantes, d'autres fois contradictoires, mais elles sont souvent complémentaires. Ces ressources externes sont accessibles via des services Web permettant de les interroger grâce à leurs API.

## 3. Appariement automatique des données

Pour faire l'appariement de ces données hétérogènes notre démarche consiste à utiliser les toponymes comme éléments d'appariement. L'appariement est divisé en 3 étapes : annotation des traces GPS, identification des différentes ENSE dans les textes de randonnées, mise en correspondance des informations extraites.

### 3.1. Annotation de la trace GPS

La technique dite du géocodage inverse permet, à partir de gazetiers, d'attribuer une adresse ou un toponyme à des coordonnées géographiques (points de la trace). Les ressources géographiques interrogées, via une paramétrisation adaptée (exemple dans le cas de l'API OSM : Zoom = 18<sup>3</sup>, Namedetails = 1<sup>4</sup>), renvoient le toponyme le plus proche selon certains critères.

---

2. GPX (eXchange Format) est un format de fichiers qui permet l'échange de coordonnées entre les dispositifs GPS ou des applications de visualisation de traces GPS. Les fichiers GPX se présentent sous la forme de fichiers XML avec une syntaxe contrôlée par un modèle.

3. Paramètre qui permet de définir le niveau de détail requis, la valeur 0 correspond au niveau pays, tandis que la valeur 18 correspond au niveau maison/bâtiment.

4. Permet d'inclure une liste de noms alternatifs dans les résultats. Celle-ci peut inclure des variantes linguistiques, et des abréviations.

Dans notre contexte pour faire du géocodage inverse il n'existe pas, à notre connaissance, de ressource géographique faisant référence. Pour l'annotation de nos traces nous avons utilisé quatre ressources complémentaires : Geonames<sup>5</sup>, OpenStreetMap (OSM)<sup>6</sup>, GoogleMaps<sup>7</sup>, Base Adresse Nationale<sup>8</sup>. Contrairement aux gazetiers (Geonames, OSM, Google) qui sont internationaux, la Base Adresse Nationale est comme son nom l'indique, une base de données des adresses postales françaises.

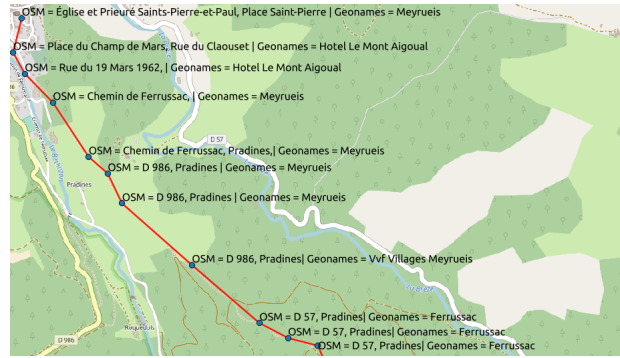


FIGURE 1. Illustration d'une trace GPS annotée

La figure 1 montre le résultat de l'annotation d'une trace GPS. Chaque point est annoté avec le résultat de son géocodage inverse. Par souci de lisibilité nous n'avons affiché que le résultat issu des ressources OSM et Geonames.

### 3.2. Geoparsing des ENSE à partir des textes de randonnées

À partir des textes nous cherchons à extraire les lieux (toponymes) et points de passage (ex : Place AlberHuile, La Forge, col de la Madeleine), appelés entités nommées spatiales étendues (ENSE) (Gaio, Moncla, 2017). Le *geoparsing* est la tâche qui permet d'extraire des ENSE dans des textes. Dans la littérature on retrouve plusieurs outils de *geoparsing* (Geoparseurs). Pour les langues romanes (français, espagnol, italien), (Moncla *et al.*, 2016) proposent un système appelé PERDIDO qui permet de reconstruire automatiquement des itinéraires à partir de textes décrivant des randonnées, le système se base sur : une cascade de transducteurs pour la phase d'annotation (*geotagging*), des ressources externes pour la phase de *geocoding*, et une méthode de clustering pour la

5. <http://api.geonames.org/findNearbyJSON>.

6. <http://locationiq.org/v1/reverse.php>

7. <https://maps.googleapis.com/maps/api/geocode/>.

8. <https://api-adresse.data.gouv.fr/reverse/>.

phase de désambiguïsation. PERDIDO a été conçu et évalué sur des textes de descriptions d'itinéraires en français, c'est principalement pour cette raison que nous avons choisi d'utiliser ce système<sup>9</sup> pour l'extraction des ENSE. Le système PERDIDO retourne un texte où les ENSE, du texte, sont annotées. Quand cela est possible, le système associe également une géolocalisation à chaque ENSE.

### 3.3. Mise en correspondance des toponymes

Après avoir annoté les traces et extrait les ENSE des textes correspondants, nous passons à la phase d'appariement. Nous présentons ici une première expérimentation d'appariement permettant de faire une évaluation préliminaire du potentiel de l'approche. Nous avons décidé de faire un appariement ponctuel, qui consiste à attribuer à chaque ENSE extraite du texte un ou plusieurs points sur la trace GPS. Nous comparons de manière récursive les ENSE de différents niveaux du texte et les toponymes de la trace annotée, en commençant par l'EN de plus bas niveau ( $EN_0$ ). Nous choisissons ensuite l'ENSE de niveau le plus haut qui a une correspondance exacte avec un toponyme issu du géocodage inverse de la trace (algorithme 1).

## 4. Résultats de l'appariement

L'expérimentation préliminaire a été réalisée sur un corpus de 8 descriptions de randonnées issues de VisoRando<sup>10</sup> chacune accompagnée de la trace GPS associée. Le tableau 1 montre pour chaque texte du corpus le nombre d'ENSE dans le texte qui correspondent à des toponymes issus du géocodage inverse des points de la trace associée.

En analysant ces résultats, on remarque que sur 115 ENSE reconnues par PERDIDO, 61 trouvent une correspondance avec les toponymes issus du géocodage inverse de la trace, ce qui fait un rappel de 53%.

Ces scores montrent bien la pertinence de notre approche qui permet de détecter à ce stade la moitié des toponymes. Les échecs de mise en correspondance sont la conséquence du non traitement de deux cas spécifiques d'entités spatiales : les premiers sont les lieux-dits et les microtoponymes absents des ressources géographiques, les seconds sont les ENSE construites uniquement à partir d'un nom commun tel que (tourner à droite de l'église, marcher jusqu'à l'auberge) qui ont une localisation implicite et relative au contexte spatial de la randonnée.

9. <http://erig.univ-pau.fr/PERDIDO/>

10. <https://www.visorando.com/>

**Algorithme 1** : Algorithme d'appariement

---

```

Data :  $L_{ENSE}, L_{GPS}$  ;
/*  $L_{ENSE}$  : Liste des ENSE extraite du texte avec Perdido
*/
/*  $L_{GPS}$  : La liste des points GPS de la trace. */
Result :  $L_{ENSE}, L_{GPS}$  ;
1 initialization;
2  $i = 0$  ;
3  $j = 0$  ;
4 while ( $j \leq nb_{ENSE}$ )  $\wedge$  ( $costPath_j \leq 0$ ) do
5   while  $j \leq nb_{GPS}$  do
6      $match = false$  ;
7      $k = 0$  ;
8     /*  $K$  : le niveau de l'ENSE */
9     while  $not(match) \parallel k \leq nb_{ENSE}$  do
10    if ( $Equal(L_{ENSE}[i][k], L_{GPS}[j])$ ) /*  $Equal$  : Une méthode
        permettant de vérifier l'égalité stricte de
        deux chaînes de caractères */
11    {  $match = true$  ;
12       $i++$  ;
13       $j++$  ;
14    }
15    else {  $k++$  ;
16    }

```

---

TABLE 1. Appariement trace GPS/ textes de randonnées

Titre de la randonnée	Nombre d'ENSE dans la randonnée	Nombre d'ENSE trouvées dans la trace
Belvédère du mont grêle	11	2
Au pays Risle	6	2
Autour de Cussac	7	7
Fontaine de Vaucluse	16	13
Au Pays de C unlhat	36	19
Sur les crêtes du Doubs	17	9
Autour de Jongieux	11	5
Le Barroux de Malaucène	11	4

**5. Conclusion**

Dans cet article, nous avons proposé une première méthode pour l'appariement de textes de randonnées et de traces associées. La méthode d'appariement proposée doit être raffinée pour traiter les différences d'écritures entre les ENSE dans le texte et les toponymes dans la trace. Une amélioration possible

serait d'introduire une mesure de similarité comme proposée par (Nguyen *et al.*, 2013). Nous comptons par la suite passer à un autre niveau d'appariement qui est l'appariement des relations spatiales entre les ENSE extraites du texte, avec des fragments de la trace GPS. À terme notre méthode devra être capable de prendre en compte les ENSE basées sur un nom commun ou celles dont le toponyme n'est pas répertorié dans les ressources géographiques utilisées.

## Bibliographie

- Buard E., Devogele T., Ducruet C. (2015). Trajectoires d'objets mobiles dans un espace support fixe. *Revue Internationale de Géomatique*, vol. 25, n° 3, p. 331–354.
- Gaio M., Moncla L. (2017). Extended named entity recognition using finite-state transducers: An application to place names. In *The ninth international conference on advanced geographic information systems, applications, and services (GEOProcessing 2017)*, p. 15–20. Nice, France, IARIA. (<hal-01492994>)
- Moncla L., Gaio M., Nogueras-Iso J., Mustière S. (2016). Reconstruction of itineraries from annotated text with an informed spanning tree algorithm. *International Journal of Geographical Information Science*, vol. 30, n° 6, p. 1137–1160.
- Nguyen V. T., Sallaberry C., Gaio M. (2013). Mesure de la similarité entre termes et labels de concepts ontologiques. In *Coria 2013*, p. 415–430. Neuchâtel, Suisse. (<hal-00847528>)
- Olteanu-Raimond A.-M., Davoine P.-A., Gaio M., Gouardères E., Van Damme M.-D., Villanova-Oliver M. *et al.* (2017, novembre). Projet CHOUCAS : Intégration de données hétérogènes et raisonnement spatial pour l'aide à la localisation des victimes en montagne. In *Spatial Analysis and GEomatics 2017 (Sagéo 20017)*. Rouen, France. (<hal-01649156>)

---

# Une approche sémantique pour représenter l'indice de végétation d'images Sentinel-2 et son évolution

Jordane Dorne<sup>1,2</sup>, Nathalie Aussenac-Gilles<sup>1</sup>,  
Catherine Comparot<sup>1</sup>, Romain Hugues<sup>2</sup>,  
Jean-Guy Planes<sup>2</sup>, Cassia Trojahn<sup>1</sup>

<sup>1</sup>*Institut de Recherche en Informatique de Toulouse, Toulouse, France*

<sup>2</sup>*Thales Alenia Space, Toulouse, France*

*{prenom.nom}@irit.fr, {prenom.nom}@thalesaleniaspace.com*

---

*RÉSUMÉ. Cet article propose une représentation sémantique (en RDF) d'indices de végétation calculés sur des images Sentinel-2. Cette représentation s'appuie sur une ontologie qui intègre la notion de tuile définie par l'ESA, i.e. une zone fixe de la surface terrestre. A chaque tuile, nous associons des indices de végétation calculés à partir des images dont l'emprise au sol correspond. Nous présentons cette ontologie et montrons comment ces représentations permettent d'étudier l'évolution de ces indices à partir de requêtes sur les données RDF.*

*ABSTRACT. In this paper we propose an RDF semantic representation of vegetation indices calculated from Sentinel-2 images. This representation relies on an ontology where the notion of tile, as defined by ESA, represents a fixed area on the Earth surface. Each tile is associated vegetation indices calculated from linked images. We present this ontology and illustrate how these representations can support the study of the evolution of these indices over time, by querying the RDF dataset.*

*MOTS-CLÉS : modèle sémantique, images satellites, indice de végétation, évolution*

*KEYWORDS: semantic model, satellite images, vegetation index, evolution*

---

## 1. Introduction

Pour collecter des données civiles d'observation de la Terre, le programme Européen Copernicus lancé en 1998 par l'Agence Spatiale Européenne (ESA) a développé une série de satellites appelés Sentinel, notamment :

- Sentinel-1 : fourniture d'imagerie radar tout-temps, jour et nuit, à des fins d'observation du sol et des océans,

SAGEO'2018 - Montpellier, 7 au 9 novembre 2018



– Sentinel-2 : fourniture d'imagerie optique haute résolution pour l'observation des sols (utilisation des sols, végétation, zones côtières, fleuves, etc.).

La grande majorité des données du programme Copernicus sont disponibles et accessibles gratuitement par tous<sup>1</sup>. Leur disponibilité ainsi que les nombreuses autres sources de données géolocalisées accessibles via le web ouvre de nombreuses perspectives économiques grâce à de nouvelles applications dans des domaines variés. Ces applications peuvent en effet bénéficier à la fois des méta-données des images (telles que la couverture nuageuse), d'une analyse automatique du contenu des images (par exemple pour y repérer la végétation, l'urbanisme ou d'autres objets d'intérêt) et de données ouvertes géo-localisées et datées pouvant être "situées" sur ces images (données gouvernementales, météorologiques, etc.). Pour associer aux images ces trois types de données (fournies et extraites des images ou collectées sur le web), l'apport des technologies sémantiques pour faciliter cette tâche a été démontré dans des travaux antérieurs (Reitsma, Albrecht, 2005; Sukhobok *et al.*, 2017), en particulier grâce à l'utilisation d'ontologies comme représentations formelles des connaissances d'un domaine donné. Ainsi, dans la continuité des travaux sur l'accès et l'intégration de données via les ontologies (Lenzerini, 2011; Lefrançois *et al.*, 2017; Console, Lenzerini, 2014), nous avons conçu un vocabulaire sémantique permettant de représenter les données des différentes sources envisagées et d'y accéder de façon homogène. Ce type d'approche s'appuie sur une ou plusieurs ontologies pour représenter les données, les relier entre elles et à l'image, et cela à un bon niveau d'abstraction. Deux problèmes se posent alors : définir une ontologie adaptée et gérer le grand volume de données ainsi générées.

Nous nous intéressons à la représentation sémantique de données extraites des images qui sont donc géo-localisées et qui évoluent dans le temps. Nous utilisons une ontologie réutilisant plusieurs vocabulaires du *Linked Open Data* reconnus comme des standards pour faciliter la réutilisation des données ainsi représentées, dont GeoSPARQL (Kolas *et al.*, 2013) pour la dimension spatiale, SOSA<sup>2</sup> pour représenter une partie des méta-données d'images satellites considérées ici comme données de capteurs, et OWL-Time<sup>3</sup> pour la dimension temporelle. Nous supposons la surface terrestre quadrillée selon une grille formée de tuiles, telle que celle fournie par l'ESA pour les images Sentinel 2 Single-Title (S2ST).

Dans cet article, nous proposons une approche pour représenter sémantiquement des indices de végétation d'images Sentinel-2. Nous représentons la notion de tuile S2ST pour associer les données directement aux tuiles. Les indices de végétation calculés à partir de ces images peuvent donc être localisés sur une tuile et datés de la date de l'image. Nous montrons comment la repré-

1. <http://copernicus.eu/data-access>

2. [https://www.w3.org/2015/spatial/wiki/SOSA\\_Ontology](https://www.w3.org/2015/spatial/wiki/SOSA_Ontology)

3. [https://www.w3.org/TR/owl-time/\(10/2017\)](https://www.w3.org/TR/owl-time/(10/2017))

sensation sémantique de ces données et leur interrogation permet d'étudier leur évolution, et ainsi l'évolution de la végétation au sol.

## 2. Ontologie proposée

L'ontologie proposée reprend celle présentée dans (Arenas *et al.*, 2018), définie pour intégrer des données géolocalisées (i.e. possédant une "géométrie") et datées autour d'images satellites. L'ontologie comporte plusieurs modules, chacun correspondant à un vocabulaire ayant un espace de nom spécifique. Le nom du module est exploité dans ce qui suit pour préfixer le vocabulaire qu'il définit. Nous présentons dans un premier temps les modules de l'ontologie réutilisée (Arenas *et al.*, 2018), c'est-à-dire les modules *sosa*, *geo*, *grid*, *eom*, puis dans un deuxième temps celui que nous avons ajouté, le module *ndvi*.

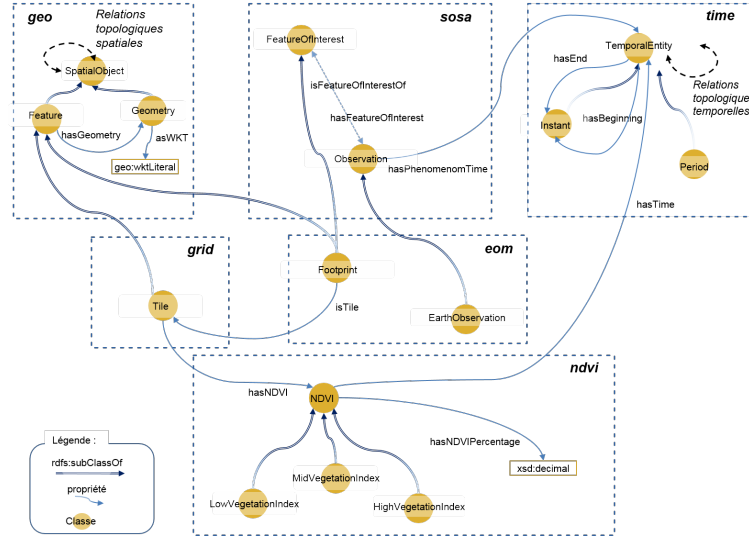


FIGURE 1. Ontologie modulaire proposée.

Le module *sosa* permet de représenter toute observation (*sosa:Observation*) réalisée par un capteur sur une "entité" (*sosa:FeatureOfInterest*), comme une activité datée (*sosa:phenomenonTime*). Le module *geo* sert à représenter la dimension spatiale des données à intégrer et réutilise GeoSPARQL. Il en reprend la classe *geo:Feature* pour représenter toute entité ou donnée géo-localisée à laquelle est associée une *geo:Geometry*, et des relations spatiales (*se trouve dans*, *recoupe*, etc.) entre les géométries. Nous utilisons ces relations pour lier deux ressources géo-localisées. Le module *grid* décrit les tuiles de l'ESA. Chaque tuile est représentée comme une instance de la classe *grid:Tile*, spécialisation de *geo:Feature* dont la propriété *geo:hasGeometry* fournit l'emprise au sol sous forme d'un polygone fermé - un carré - défini par des coordonnées GPS. Le module *eom* sert à représenter les méta-données d'images (module simplifié à deux classes sur la Figure 1). La propriété *eom:isTile* permet d'associer

une tuile au `eom:Footprint` de chacune des images S2ST lui correspondant : un footprint représente la zone géographique couverte par une image et est décrit par un polygone fermé (une géométrie). Ce footprint est considéré par ailleurs comme le point d'intérêt d'observations réalisées par des capteurs du satellite ayant pris l'image ; la classe `eom:Footprint` spécialise ainsi `geo:Feature` et `sosa:FeatureOfInterest`. Le `eom:Footprint` d'une image est ainsi daté via la propriété `sosa:phenomenonTime` qui identifie le moment où l'image a été prise.

Adapter l'ontologie à la représentation d'un nouveau type de données (comme les indices de végétation) revient à ajouter à ces modules un nouveau module pour ces données, le module *ndvi*, et à le lier aux modules existants. Chaque cadre de la Figure 1 présente un extrait du vocabulaire d'un module exploité pour associer les indices NDVI aux images S2ST. Au sein du module *ndvi*, l'indice de végétation est représenté par la classe `ndvi:NDVI`. Les propriétés spatiale et temporelle des indices de végétation sont fournies respectivement par la tuile à laquelle ils sont associés et la relation `time:hasTime` (OWL Time). Cette dernière permet d'exprimer des relations temporelles (*avant*, *après*, etc.) entre deux entités temporelles.

### 3. Cas d'étude

L'indice de végétation NDVI (*Normalized Difference Vegetation Index*) est calculé à l'aide des images S2ST de niveau L1C. Nous avons choisi les images comportant une couverture nuageuse inférieure à 3% afin de ne pas fausser les résultats. Cet indice est obtenu par un calcul sur les bandes "proche infrarouge" (PIR) et "rouge" (R) du capteur, qui permet de détecter le niveau de chlorophylle. Le résultat de cette opération donne une matrice de valeurs comprises entre -1 et 1 caractérisant le NDVI de chaque pixel de l'image. Les valeurs comprises entre -1 et 0 représentent les éléments composés d'eau, les valeurs comprises entre 0 et 0.25 représentent les éléments composés de terre. Nous nous intéressons ici uniquement aux valeurs concernant la végétation, c'est à dire aux valeurs comprises entre 0.25 et 1. Plus l'indice est élevé plus la végétation est dense. Pour associer un indice à chaque pixel d'une image, nous avons classé les indices en trois catégories : *LowVegetationIndex* (NDVI compris entre 0.25 et 0.5), *MidVegetationIndex* (NDVI compris entre 0.5 et 0.75), et *HighVegetationIndex* (NDVI compris entre 0.75 et 1), correspondant chacune approximativement à de l'herbe, de la culture, et de la forêt.

Pour obtenir une représentation de ces catégories en pourcentage d'image il faut calculer le nombre de pixels de chaque catégorie par rapport au nombre total de pixels de l'image. Nous avons représenté ces trois catégories dans le module *ndvi* sous forme de classes (Cf. Figure 1). Les indices de végétation calculés, instances d'une de ces classes, sont datés à partir de la date de prise de l'image et géo-localisés avec la tuile associée à l'image. Ci-après un exemple de graphe RDF représentant, selon notre modèle, l'indice de végétation Low (instance de `ndvi:LowVegetationIndex`) de la tuile 31TCJ au 21 avril 2018.

```

g-ndvi:ndvi_31TCJ_20180421T105031_low a ndvi:LowVegetationIndex .
g-grid:31TCJ a grid:Tile .
g-grid:31TCJ ndvi:hasNdvi g-ndvi:ndvi_31TCJ_20180421T105031_low .
g-ndvi:ndvi_31TCJ_20180421T105031_low
  ndvi:hasNdviPercentage "15.11"^^xsd:decimal .
g-ndvi:ndvi_31TCJ_20180421T105031_low
  time:hasTime ndvi:instant_1524300631 .
ndvi:instant_1524300631 a time:Instant .
ndvi:instant_1524300631
  time:inXSDDateTime "2018-04-21T10:50:31.026000"^^xsd:dateTime .
ndvi:instant_1524300631
  time:inXSDDateTimeStamp "1524300631"^^xsd:dateTimeStamp .

```

time	tileId	ndviHighPercent	ndviMidPercent	ndviLowPercent
2017-05-26T10:50:31.026	<a href="http://melodi.irit.fr/iod/grid_S2ST/31TCJ">http://melodi.irit.fr/iod/grid_S2ST/31TCJ</a>	18.22	47.09	16.91
2017-07-05T10:50:31.026	<a href="http://melodi.irit.fr/iod/grid_S2ST/31TCJ">http://melodi.irit.fr/iod/grid_S2ST/31TCJ</a>	16.11	35.22	23.81
2017-08-14T10:50:31.026	<a href="http://melodi.irit.fr/iod/grid_S2ST/31TCJ">http://melodi.irit.fr/iod/grid_S2ST/31TCJ</a>	8.19	36.38	31.73
2017-10-08T10:50:09.027	<a href="http://melodi.irit.fr/iod/grid_S2ST/31TCJ">http://melodi.irit.fr/iod/grid_S2ST/31TCJ</a>	2.66	37.25	23.83
2017-10-13T10:50:31.026	<a href="http://melodi.irit.fr/iod/grid_S2ST/31TCJ">http://melodi.irit.fr/iod/grid_S2ST/31TCJ</a>	0.45	35	26.2
2017-10-28T10:51:29.027	<a href="http://melodi.irit.fr/iod/grid_S2ST/31TCJ">http://melodi.irit.fr/iod/grid_S2ST/31TCJ</a>	0.38	31.12	26.38
2017-11-07T10:52:29.027	<a href="http://melodi.irit.fr/iod/grid_S2ST/31TCJ">http://melodi.irit.fr/iod/grid_S2ST/31TCJ</a>	0.33	29.3	28.77
2017-11-22T10:53:41.026	<a href="http://melodi.irit.fr/iod/grid_S2ST/31TCJ">http://melodi.irit.fr/iod/grid_S2ST/31TCJ</a>	0.16	14.7	38.44
2017-11-27T10:53:59.027	<a href="http://melodi.irit.fr/iod/grid_S2ST/31TCJ">http://melodi.irit.fr/iod/grid_S2ST/31TCJ</a>	0.23	10.77	43.09
2018-02-10T10:52:01.026	<a href="http://melodi.irit.fr/iod/grid_S2ST/31TCJ">http://melodi.irit.fr/iod/grid_S2ST/31TCJ</a>	0.44	18.17	55.53
2018-02-25T10:50:19.027	<a href="http://melodi.irit.fr/iod/grid_S2ST/31TCJ">http://melodi.irit.fr/iod/grid_S2ST/31TCJ</a>	0.24	15.84	58.84
2018-04-21T10:50:31.026	<a href="http://melodi.irit.fr/iod/grid_S2ST/31TCJ">http://melodi.irit.fr/iod/grid_S2ST/31TCJ</a>	12.48	54.44	15.11

FIGURE 2. Résultat SPARQL du NDVI sur une année pour la tuile 31TCJ.

Un graphe de ce type est généré pour chaque catégorie d'indice d'une image à partir d'un script Python qui prend en entrée le fichier jpeg2000 de l'image, calcule les NDVI correspondants et peuple l'ontologie. Sur la Figure 2, le résultat d'une requête SPARQL indique l'évolution du NDVI de la tuile 31TCJ sur une année (filtrage sur la composante temporelle des NDVI). Grâce à ces valeurs, nous pouvons observer l'évolution de la végétation en fonction des saisons. Sur cet exemple, nous constatons au début du printemps (entre fin février et fin mai) une augmentation du pourcentage de la catégorie *HighVegetationIndex*, correspondant à la croissance du feuillage de forêts ; à l'inverse, nous notons une diminution de ce pourcentage pendant la période hivernale. Les catégories *MidVegetationIndex* et *LowVegetationIndex* sont aussi impactées par le changement de saisons.

#### 4. Conclusion

Dans cet article, nous avons proposé de représenter via une ontologie les indices de végétation des tuiles Sentinel-2 calculés à partir des images qui leur sont liées. L'ontologie sert aussi à établir des relations temporelles permettant l'analyse de l'évolution des indices. Cela peut être utile aux analyses en agriculture. Ce travail complète celui que nous avons mené pour l'intégration de données contextuelles aux métadonnées d'images via des relations spatio-temporelles (Arenas *et al.*, 2018). Il s'agit d'une première étape en vue d'identifier à terme un changement anormal. Nous envisageons de fournir des règles et des fonctionnalités de raisonnement pour faciliter les analyses.

#### Remerciements

Ce travail est co-financé par le projet CANDELA (convention de subvention num 776193 du programme de recherche et d'innovation Horizon 2020 de l'Union européenne), et par la convention Cifre ANR n° 2017/1399 entre TAS et le CNRS.

#### Bibliographie

- Arenas H., Aussenac-Gilles N., Comparot C., Trojahn C. (2018). Ontologie pour l'intégration de données d'observation de la Terre et contextuelles basée sur les relations topologiques. In *29es Journées Francophones d'Ingénierie des Connaissances*, p. 5-20. Nancy, France, AFIA.
- Console M., Lenzerini M. (2014). Reducing global consistency to local consistency in ontology-based data access - extended abstract. In *Informal Proceedings of the 27th International Workshop on Description Logics*, p. 496-499. Vienna, Austria.
- Kolas D., Perry M., Herring J. (2013). *Getting started with GeoSPARQL*. Rapport technique. OGC. Consulté sur [http://www.ssec.wisc.edu/meetings/geosp\\_sem/presentations/GeoSPARQL\\_Getting\\_Started-KolasWorkshopVersion.pdf](http://www.ssec.wisc.edu/meetings/geosp_sem/presentations/GeoSPARQL_Getting_Started-KolasWorkshopVersion.pdf)
- Lefrançois M., Zimmermann A., Bakerally N. (2017, mai). A SPARQL extension for generating RDF from heterogeneous formats. In E. Blomqvist, D. Maynard, A. Gangemi, R. Hoekstra, P. Hitzler, O. Hartig (Eds.), *Proc. of the 2017 Extended Semantic Web Conference, Part I*, p. 35-50. Portoroz, Slovenia, Springer.
- Lenzerini M. (2011). Ontology-based data management. In *Proceedings of the 20th acm international conference on information and knowledge management (cikm '11)*, p. 5-6. New York, NY, USA, ACM.
- Reitsma F., Albrecht J. (2005). Modeling with the semantic web in the geosciences. *IEEE Intelligent Systems*, vol. 20, n° 2, p. 86-88.
- Sukhobok D., Sánchez H., Estrada J., Roman D. (2017). Linked data for common agriculture policy: Enabling semantic querying over sentinel-2 and lidar data. In N. Nikitina, D. Song, A. Fokoue, P. Haase (Eds.), *Iswc 2017 posters, demonstrations and industry tracks (iswc-pd-industry)*. Aachen. Consulté sur <http://ceur-ws.org/Vol-1963/#paper559>

# Estimation de la hauteur de la végétation à partir des ombres portées des images Très Haute Résolution Spatiale

Anne-Sophie Onody<sup>1</sup>, Fatoumata Bah<sup>2</sup>, Grégory Duval<sup>2</sup>, Pauline Dusseux<sup>1</sup>

1. Agrocampus Ouest  
65 rue de Saint Brieuc, 35042 Rennes Cedex, France  
anne-sophie.onody@agrocampus-ouest.fr

2. SNCF Réseau, Direction Ingénierie et Projets (I&P LVE ENV)

---

*RESUME.* Dans le cadre des missions d'entretien, SNCF Réseau cherche à évaluer l'utilisation de l'imagerie satellitaire pour distinguer les différentes strates de végétation présentes dans les emprises ferroviaires. Les recherches présentées ont pour objectif d'estimer la hauteur de la végétation à partir des ombres portées d'images Pléiades à Très Haute Résolution Spatiale. Les résultats obtenus sont comparés à des hauteurs mesurées sur le terrain et à un modèle numérique de surface normalisé.

*ABSTRACT.* In the context of the maintenance missions, SNCF Réseau seeks to evaluate the use of satellite imagery to distinguish the different vegetation strata on the railway rights-of-way. The objective of these researches is to estimate the vegetation height from shadow analysis on Pleiades Very High-Resolution Satellite Imagery. The results are compared to real heights and to a standardized digital surface model.

*MOTS-CLÉS :* télédétection, ombre, hauteur de végétation

*KEYWORDS:* remote sensing, shadow analysis, vegetation height

---

## 1. Introduction

SNCF Réseau, le gestionnaire du réseau ferré français, est en charge de la maintenance des voies ferrées et de leurs abords. La maîtrise de la végétation est un des enjeux forts de cette maintenance et une nécessité aussi bien pour la sécurité que pour la mobilité des trains et des agents. Au vu de l'importance du réseau sur le territoire français, la recherche d'une maîtrise de la végétation au coût de revient le plus bas possible, tout en garantissant la sécurité et la pérennité des installations dans le respect de l'environnement, est primordiale (cf SNCF Réseau, 2012).

Les recherches en cours ont ainsi pour objectif d'évaluer l'utilisation de l'imagerie satellitaire pour distinguer les différentes strates végétales présentes dans les emprises des voies ferrées pour planifier les interventions d'entretien et de modernisation

SAGEO'2018 – Montpellier, 6-9 novembre 2018

nécessaires. La présente étude se concentre sur l'exploitation des ombres portées pour estimer la hauteur de la végétation et classer la végétation selon une typologie en trois classes : strate herbacée (hauteur inférieure à 1m) strate arbustive (hauteur comprise entre 1m et 6m), et strate arborée (hauteur supérieure à 6m). Les résultats obtenus sont ensuite comparés aux hauteurs mesurées sur le terrain en avril 2018 et à un modèle numérique de surface normalisé calculé à partir d'une paire d'images stéréo Pléiades.

## 2. Méthodologie

### 2.1. Les données utilisées

L'image satellitaire utilisée est une image Très Haute Résolution Spatiale (THRS) stéréo Pléiades de résolution spatiale de 0.5m dans le panchromatique acquise sur la région lilloise et datant du 14 février 2017. L'image est reprojetée en Lambert 93 et les traitements sont réalisés sur une zone d'étude d'une surface d'environ 150 ha le long de la LGV Nord aux alentours de la commune de Prêmesques à l'Est de Lille. Un modèle numérique de terrain (MNT) d'une résolution de 5m réalisé par l'IGN est également utilisé. Des mesures ont été effectuées sur le terrain à l'aide d'un théodolite. Peu de données ont pu être récoltées car les emprises ferroviaires de la LGV sont particulièrement difficiles d'accès pour des raisons de sécurité.

### 2.2. Calcul d'un modèle numérique d'élévation et étude de la topographie sur la zone d'étude

Un modèle numérique de surface (DSM) est calculé à l'aide du logiciel s2p, un outil libre et gratuit permettant une reconstruction 3D automatique à partir d'images stéréo à haute résolution spatiale (cf de Franchis *et al.*, 2014). Un modèle numérique de surface normalisé (nDSM) en est ensuite extrait par soustraction du MNT. Parallèlement le MNT est utilisé pour calculer la pente et l'exposition de la zone d'étude à l'aide du logiciel QGIS.

### 2.3. Extraction des ombres et estimation de la hauteur des objets

Un seuillage sur les valeurs numériques de la bande panchromatique (Figure 1.a.) est réalisé afin de distinguer les pixels correspondants aux ombres des autres pixels (Figure 1.b.) d'après la méthode proposée par Dare (2005). Des boîtes englobantes d'emprise orientée minimale sont ensuite construites autour des différentes zones d'ombre (Figure 1.c.) et les valeurs de longueur, largeur et orientation des rectangles obtenus en sont extraites. La combinaison de ces méthodes permet d'extraire les zones d'ombre et de calculer leur longueur selon la direction du Nord puis selon la direction de l'azimut du soleil comme détaillé dans la Figure 1.d.

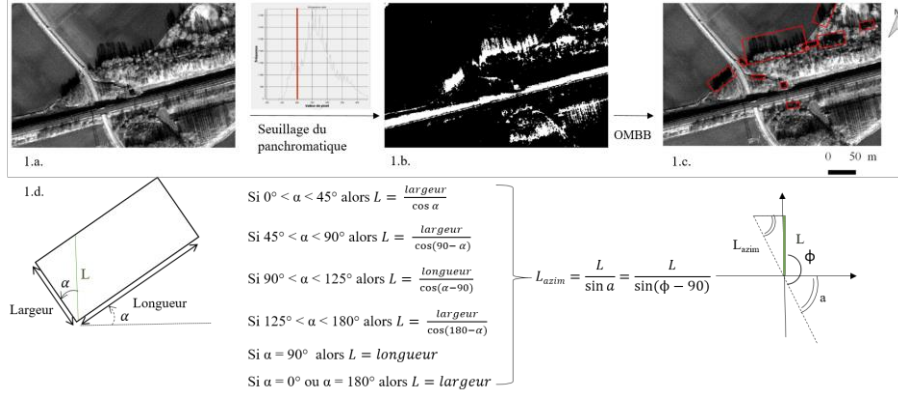


FIGURE 1. Extrait de l'image panchromatique de la zone d'étude (1.a.), du masque « ombre » obtenu par seuillage (1.b.), des exemples de boîtes englobantes obtenues à l'aide de l'outil OMBB de QGIS (1.c.) et le calcul de la longueur de l'ombre selon la direction du Nord ( $L$ ) et la direction de l'azimut solaire ( $L_{\text{azim}}$ ) (1.d.)

Le calcul de statistiques de zones permet d'obtenir la pente moyenne et l'exposition dans chacun des rectangles et de connaître ainsi la topographie de la zone. La longueur de l'ombre portée et l'estimation de la hauteur des différents objets sont ensuite calculés selon les différentes configurations (Figure 2) (cf Cheng et Thiel, 1995). Les mêmes calculs sont réalisés avec  $L_{\text{azim}}$ , ce qui permet d'obtenir quatre valeurs différentes d'estimation de la hauteur (selon la longueur choisie et la prise en compte ou non de la topographie).

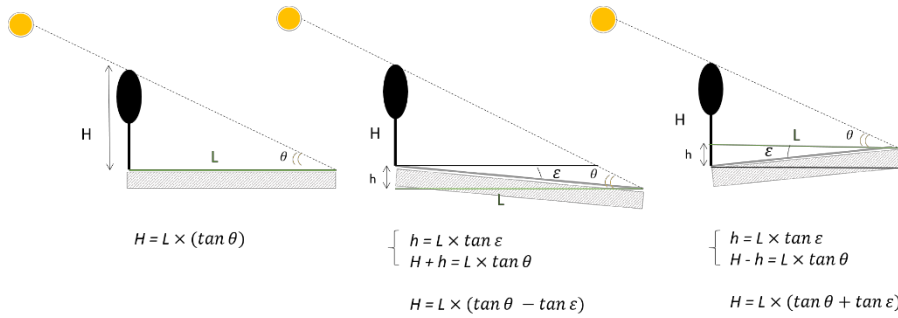


FIGURE 2. Relations trigonométriques entre la longueur de l'ombre mesurée et la hauteur de la végétation selon la topographie.  $H$  = hauteur de la végétation,  $L$  = longueur de l'ombre mesurée sur l'image,  $\theta$  = l'élévation du soleil,  $\epsilon$  = la pente du terrain.



### 3. Résultats

Dix-neuf points de contrôle correspondant à des bâtis ou à de la végétation sont sélectionnés sur l'ensemble de la zone d'étude (Figure 3). Une interprétation visuelle préalable permet de sélectionner une vingtaine de boîtes englobantes utilisées pour comparer les estimations de hauteurs calculées et les valeurs d'élévation extraites du nDSM. Des vérités terrain ont pu être récoltées pour certains de ces points de contrôle et permettent de constater la relative précision du nDSM (Table 1). La méthodologie présentée semble très légèrement surestimer les hauteurs.

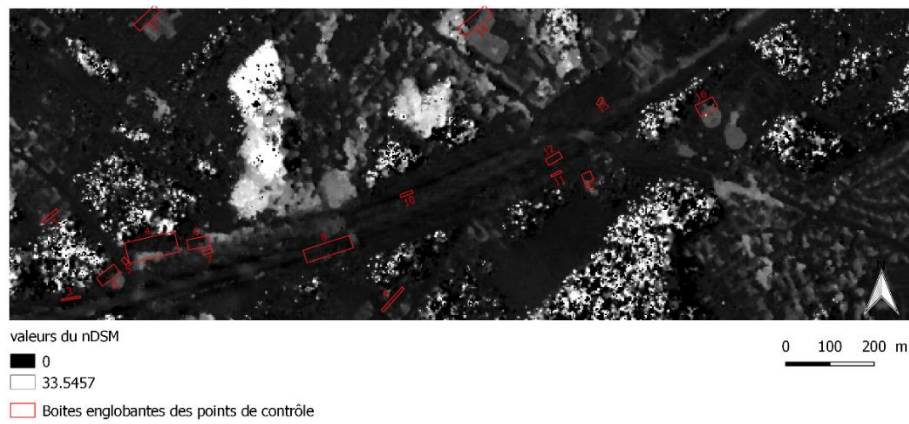


FIGURE 3. *nDSM de la zone d'étude et localisation des boîtes englobantes des points de contrôle.*

TABLE 1. Comparaison des hauteurs estimées, des valeurs du nDSMet des données terrains pour les 19 points de contrôle détaillés dans la Figure 3

Point s	H <sub>1</sub>	H <sub>2</sub>	H <sub>3</sub>	H <sub>4</sub>	nDS M	Terrai n	Erreur 1	Erreur 2	Erreur 3	Erreur 4
1	4.42	4.42	3.79	3.79	3.6	/	+0.82	+0.82	+0.19	+0.19
2	14.0 6	11.2 4	12.2 2	9.76	9.5	8.5	+4.56	+1.74	+2.72	+0.26
3	1.47	0.82	1.45	0.81	0	1.5	+1.47	+0.82	+1.45	+0.81
4	23.3 1	19.7 2	22.7 1	19.2 1	15.5	13.5	+7.81	+4.22	+7.21	+3.71
5	6.31	5.70	8.61	7.78	7	/	-0.69	-1.3	+1.61	+0.78
6	11.2 9	7.40	11.1 3	7.29	7.6	3.8	+3.69	-0.2	+3.53	-0.31
7	4.04	2.53	3.87	2.43	3.10	/	+0.94	-0.57	+0.77	-0.67
8	14.7 3	9.44	14.0 6	9.01	4.7	4.15	+10.0 3	+4.74	+9.36	+4.31
9	3.88	3.64	5.05	4.73	4.20	/	-0.32	-0.56	+0.85	+0.53
10	3.11	2.44	2.93	2.30	1.70	/	+1.41	+0.74	+1.23	+0.60
11	2.66	2.20	2.46	2.03	2.0	/	+0.66	+0.20	+0.46	+0.03
12	8.41	3.00	7.59	2.71	2.0	/	+6.41	+1.00	+5.59	+0.71
13	4.01	3.77	4.86	4.56	4.0	/	+0.01	-0.23	+0.86	+0.56
14	17.0	15.6 4	14.6 0	13.4 2	9.5	9.50	+7.5	+6.14	+5.1	+3.92
15	10.9 3	10.2 4	10.2 6	9.61	9.5	/	+1.43	+0.74	+0.76	+0.11
16	14.6 2	13.7 8	13.4 2	12.6 5	10	/	+4.62	+3.78	+3.42	+2.65
17	12.6 2	12.7 4	10.6 1	10.7 1	10.5	/	+2.12	+2.24	+0.11	+0.21
18	9.17	8.58	8.73	8.18	8	/	+1.17	+0.58	+0.73	+0.18
19	56.4 0	54.9 5	48.7 3	47.4 8	41.8	/	+14.6	+13.1 5	+6.93	+5.68
<b>Err. moy</b>							<b>+3.59</b>	<b>+2.00</b>	<b>+2.78</b>	<b>+1.38</b>

H<sub>1</sub> : ni l'orientation du soleil ni la topographie ne sont pris en compte ; H<sub>2</sub> : prise en compte de la topographie uniquement; H<sub>3</sub> : prise en compte de l'orientation du soleil uniquement ; H<sub>4</sub> : prise en compte de l'orientation du soleil et de la topographie.

$$\text{Erreur}^i = H_i - \text{nDSM}$$

#### 4. Conclusion et perspectives

La méthode présentée ici a pour avantage d'être facilement reproductible et semi-automatisable ce qui répond aux attentes de SNCF Réseau dont la propriété foncière est de 95 000ha. Elle permet également d'obtenir une information de hauteur sans avoir recours à des images stéréo qui sont plus rares et plus coûteuses. L'erreur absolue moyenne calculée sur les points de contrôle (1.38m) est suffisamment faible

pour permettre de distinguer les trois strates végétales demandées : herbacée, arbustive et arborée.

Cependant, de même que dans la publication de Shettigara et Sumerling (1998), quelques limites peuvent être soulignées. Tout d'abord, la longueur de l'ombre projetée doit être supérieure à la taille d'un pixel, à savoir 0.5m pour une image Pléiades, ce qui correspond à une hauteur supérieure à 0.25m (en utilisant une image d'hiver ; ce seuil sera plus élevé en été à cause de l'élévation du soleil). La méthode de détection des ombres mériterait d'être affinée d'après les méthodes proposées par Dare (2005) afin d'augmenter encore la précision de la largeur de l'ombre et donc l'estimation de la hauteur. Enfin, les zones d'ombre ne doivent pas intercepter d'autres objets, ce qui risquerait d'introduire un biais en augmentant la largeur de l'ombre mesurée sur l'image et donc l'estimation de hauteur. Ainsi, cette méthode est adaptée à l'estimation de la hauteur d'alignements d'arbres ou de lisières, pouvant être d'intérêt pour répondre plus largement à la caractérisation de continuités écologiques telles que la trame boisée.

### Références bibliographiques

- Cheng F., Thiel K-H. (1995), Delimiting the building heights in a city from the shadow in a panchromatic SPOT-image Part 1. Test of forty-two buildings, *International Journal of Remote Sensing*, Vol 16 n°3, p. 409-415.
- Dare P-M (2005), Shadow Analysis in High-Resolution Satellite Imagery of urban areas, *Photogrammetric Engineering & Remote Sensing* Vol. 71, n°2, pp. 169–177.
- de Franchis C., Meinhardt-Llopis C., Michel J., Morel J.-M., Facciolo G. (2014), An automatic and modular stereo pipeline for pushbroom images, *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. II-3, p. 49-56.
- Shettigara V.K., Sumerling G.M. (1998), Height determination of extended objects using shadows in SPOT images, *Photogrammetric Engineering & Remote Sensing*, Vol. 64, n°1, p. 35-44 .
- SNCF RESEAU (2012), IN7136 Gestion des dépendances vertes, *Référentiel d'infrastructure*.

---

# Le vandalisme dans l'information géographique volontaire : apprendre pour mieux détecter ?

Quy Thy Truong<sup>1</sup>, Guillaume Touya<sup>1</sup>, Cyril de Runz<sup>2</sup>

1. Univ. Paris-Est, LASTIG COGIT, IGN, ENSG, F-94160 Saint-Mande, France

*quy-thy.truong@ign.fr; guillaume.touya@ign.fr*

2. Modeco, CReSTIC, University of Reims Champagne-Ardenne, CS 30012, Reims cedex 2, France

*cyril.de-runz@univ-reims.fr*

---

*RÉSUMÉ. Dans un contexte où le vandalisme de l'information géographique volontaire constitue une réelle menace pour la qualité des données, cet article propose d'explorer les différentes techniques permettant de le détecter. Tout d'abord, nous tâchons d'examiner les différentes définitions du vandalisme, mettant en avant la complexité de cette notion. Puis, nous expérimentons la détection du vandalisme des données OpenStreetMap (OSM) en utilisant une méthode de détection d'anomalie par classification non-supervisée. L'analyse de nos résultats initiaux conduit alors à une discussion sur la pertinence de construire un corpus de vandalisme des données OSM utilisable dans un contexte de classification supervisée.*

*ABSTRACT. As vandalism is a serious matter for the quality of Volunteered Geographic Information, this paper aims at exploring learning techniques that enable its detection. First, a focus on the various definitions of vandalism highlights the complexity of this concept. Second, we present an experimental vandalism detection on OpenStreetMap (OSM) data using a clustering-based outlier detection. The analysis of initial results leads to a discussion about the construction of an OSM vandalism corpus that would be useful in a supervised learning context.*

*MOTS-CLÉS : Information Géographique Volontaire, Vandalisme, Classification*

*KEYWORDS: Volunteered Geographic Information, Vandalism, Learning*

---

## 1. Introduction

Le manque de qualification de l'information géographique volontaire constitue aujourd'hui encore un frein pour les consommateurs de données géographiques. En effet, dans un contexte où la science citoyenne permet à des non-spécialistes de collecter des données scientifiques, la saisie collaborative d'information géographique par des volontaires questionne actuellement les producteurs de données institutionnelles quant au mode d'acquisition de l'information spatiale et la mise à jour de leurs bases de données (Ivanovic, 2018 ; Noucher, 2014). Cependant, alors que les bases de données géographiques participatives peuvent contenir des éléments de qualité variable, le fait de permettre à n'importe quel volontaire de contribuer donne la possibilité au plus mal-intentionné de réduire la qualité des données par des actes de vandalisme cartographique (Ballatore, 2014). Des outils de détection du vandalisme de données Wikimédia (Potthast, 2010 ; Heindorf *et al.*, 2015) et OpenStreetMap (Neis *et al.*, 2012) ont été développés pour prévenir ce genre de dérives, or ceux-ci n'ont pas été conçus selon la même conception de vandalisme : en effet, la frontière entre le vandalisme et la simple erreur de saisie étant très fine, certains scientifiques choisissent une définition inclusive du vandalisme qui comprend alors les actes de dégradation involontaire. La détection du vandalisme dans l'information géographique volontaire est intéressante car elle peut constituer une étape préalable à toute utilisation de ce type de données. En effet, le filtrage de données vandalisées permet de garantir une qualité minimale à un jeu de données collaboratives. Par ailleurs, les données vandalisées obtenues présentent aussi un intérêt pour l'approfondissement des recherches sur le carto-vandalisme.

Deux questions découlent de ce constat : comment définir le vandalisme dans l'information géographique volontaire et comment le détecter pour le corriger ? Dans cet article, nous dressons un état de l'art sur vandalisme afin de mettre en évidence les définitions formulées dans la littérature, et nous explorons les différentes approches de détection du vandalisme qui ont été expérimentées jusqu'à présent. Nous cherchons alors à détecter le vandalisme des données OpenStreetMap (OSM) en expérimentant une méthode de détection d'anomalie. Le but de notre expérience est de déterminer si un algorithme de détection d'anomalie par clustering est capable de retrouver des cas très simples de vandalisme artificiels, présentant des anomalies géométriques. L'interprétation de nos résultats expérimentaux ouvriront la discussion sur les pistes à envisager pour améliorer la détection.

## 2. Définitions théoriques et pratiques du vandalisme

### 2.1. Bref historique

Historiquement, le vandalisme désignait les actes de dégradation perpétrés par la tribu des Vandal, un peuple germanique réputé pour avoir pillé et sac-

cagé les oeuvres d'art et les monuments durant leur invasion de l'Occident. Pendant la période révolutionnaire, l'Abbé Grégoire reprend le terme de vandalisme pour désigner les dommages causés sur les monuments et les ouvrages d'arts du patrimoine français<sup>1</sup>. Aujourd'hui le terme de vandalisme renvoie plus généralement à une dégradation matérielle causée par des êtres humains. La justice française définit le vandalisme de la manière suivante<sup>2</sup> :

"L'acte de vandalisme consiste à détruire, dégrader, ou détériorer volontairement le bien d'autrui. (...) L'acte de vandalisme doit être commis sans motif légitime. Il est par exemple permis de briser une vitre pour sauver une personne en danger."

D'après cette définition, trois conditions doivent être remplies pour pouvoir qualifier un acte de vandalisme : qu'il y ait observation d'un bien dégradé ; que la dégradation ait été voulue par son auteur ; qu'il n'y ait aucun élément de contexte justifiant la légitimité de cet acte. En effet, selon le contexte, un même acte ne sera pas forcément labélisé de la même manière : par exemple, faire dérailler un train en déposant des obstacles sur les rails sera considéré comme une farce si cela a été fait par des enfants (en supposant qu'il n'ont pas encore conscience de l'ampleur de leurs actes), mais comme du vandalisme si cela a été commis par des adultes (Zimbardo, 1971).

Les travaux de recherche sur le vandalisme de l'information géographique volontaire ont proposé deux typologies de vandalisme cartographique. Neis *et al.* (2012) ont développé un outil de détection du vandalisme, où le vandalisme comprend tout dommage causé volontairement ou non. Cette typologie rompt avec la définition traditionnelle du vandalisme, puisqu'elle met de côté la composante intentionnalité. Ballatore (2014) propose une typologie de carto-vandalisme à partir des différentes motivations qui poussent à l'acte. Le vandalisme cartographique peut être : ludique, idéologique, fantaisiste, artistique, industriel ou spam cartographique. Bien que cette typologie soit empirique, il est difficile de l'utiliser en pratique car les différentes classes proposées ne sont pas indépendantes les unes des autres. Par exemple, le carto-vandalisme artistique peut être vu comme une sous-catégorie de carto-vandalisme fantaisiste puisque des objets cartographiés artistiquement sont également des objets fictifs.

On souligne ici que l'évaluation des trois composantes du vandalisme est ardue : elle fait appel à de véritables compétences juridiques, ce qui explique la difficulté de qualifier un acte de vandalisme. Dans le but d'appliquer cette définition du vandalisme au domaine de l'information géographique volontaire, nous présentons dans la sous-section suivante une exploration et une analyse de différents cas de contributeurs bannis sur OSM. A la lumière de la définition

1. <http://www2.assemblee-nationale.fr/decouvrir-l-assemblee/histoire/grands-moments-d-eloquence/l-abbe-gregoire-31-aout-1794>

2. <https://www.service-public.fr/particuliers/vosdroits/F1514>

précédemment dégagée, nous cherchons à qualifier les cas de contributions qui relèvent du vandalisme.

## 2.2. Exploration des contributeurs bannis sur OpenStreetMap

Sur OpenStreetMap, il existe un groupe de contributeurs – le Data Working Group<sup>3</sup> (DWG) – dont le rôle est de bannir temporairement les comptes OSM présentant des activités douteuses, notamment ceux qui vandalisent les données. Une liste des contributeurs bannis est consultable sur une page web dédiée<sup>4</sup>.

Le premier cas de vandalisme présenté dans la Figure 1 est un élément géographique correspondant initialement à une zone commerciale, qui a été transformée en lac par un second contributeur (Table 1). Ici, la dégradation causée est assez évidente : l'objet géographique a complètement changé de nature puisqu'il ne décrit plus du tout la même réalité, et la présence d'un lac sur une zone où se trouvent déjà un parking et des commerces paraît assez invraisemblable. La question de la vraisemblance des relations topologiques entre les objets géographiques a été soulevée pour identifier les problèmes de qualité dans OSM, qui peuvent relever de simples erreurs, d'incohérences de niveau de détail voire de vandalisme (Goodchild, Li, 2012 ; Touya, Brando, 2013). De plus, le fait de nommer le lac par un smiley (tag `name = :)`) révèle que le contributeur a certainement dégradé les données de manière intentionnelle. Enfin, en observant les autres contributions de ce compte utilisateur, ce dernier a ajouté de la même manière d'autres lacs en zones urbaines, ce qui laisse peu de doute quant à sa culpabilité. On peut donc conclure que cette contribution relève bien d'un acte de vandalisme.

TABLE 1. Historique des tags de l'objet de la Figure 1

Tag key	Version	1	2
landuse		retail	
name			:)
natural			water
water			lake

Le cas de la Figure 2 correspond plutôt à une erreur de débutant qu'à un pur acte de vandalisme. En effet, bien que l'hôpital soit géométriquement mal cartographié, on note que l'objet géographique a été ajouté trois heures après la date de création du compte OSM de son auteur. De plus, le commentaire laissé par le contributeur sur cette session d'édition et ses contribution suivantes

3. [https://wiki.openstreetmap.org/wiki/FR:Data\\_working\\_group](https://wiki.openstreetmap.org/wiki/FR:Data_working_group)

4. [https://www.openstreetmap.org/user\\_blocks](https://www.openstreetmap.org/user_blocks)

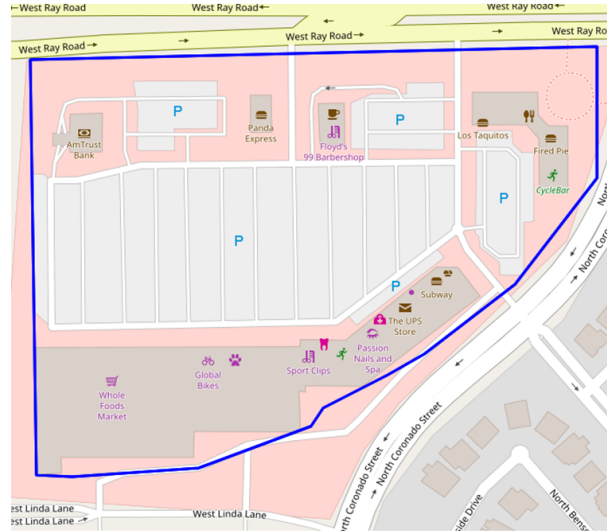


FIGURE 1. Exemple d'un cas de vandalisme : la zone commerciale est transformée en lac (cf. Table 1 pour l'historique de tags de l'objet)

indiquent que celui-ci cherchait initialement à cartographier les éléments manquants aux alentours d'une université au Sri Lanka<sup>5</sup>. Ici, il s'agit probablement d'un problème de système de coordonnées, et non d'un cas de vandalisme.

Enfin, l'historique des modifications de tags de l'objet de la Table 2 montre un désaccord entre les contributeurs sur la toponymie de la ville. Le contributeur de la version n°32 de l'objet a été banni pour avoir retiré l'orthographe serbe du tag `name`. Après avoir été banni, la version n°32 a été annulée (i.e. la version n°33 correspond à la version n°31). Au vu du grand nombre de versions de l'objet – plus de 30 – celui-ci semble être un sujet controversé pour les contributeurs OSM. Or, précisons que les guerres d'éditions ne sont pas autorisées sur OSM<sup>6</sup>, celles-ci sont même considérées comme étant du vandalisme d'après le règlement du projet. Le contributeur de la version n°32 a donc été banni pour non respect des règles de bonne conduite du projet. Cependant, même si sur la version n°32, l'orthographe serbe ne figure pas sur le tag `name`, celui-ci reste indiqué sur le tag `name:sr-Latn`. Par ailleurs, d'après l'OSCE<sup>7</sup>, la ville de Malishevë/Mališevo est habitée majoritairement par des albanais, ce

5. <https://www.openstreetmap.org/changeset/53283079>

6. <https://wiki.openstreetmap.org/wiki/Disputes>

7. <https://www.osce.org/kosovo/13137?download=true>



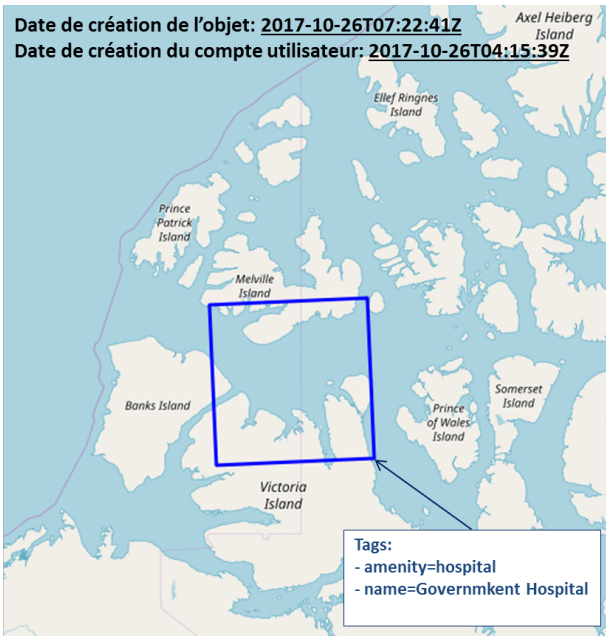


FIGURE 2. Erreur de débutant : cartographie d'un hôpital dont la géométrie chevauche plusieurs îles au Nord du Canada.

qui pourrait justifier l'acte de ce contributeur. Par conséquent, on ne peut pas affirmer que sa contribution dégrade la qualité de l'objet.

TABLE 2. Extrait de l'historique des tags (versions n°28-n°33) d'une ville du Kosovo dont la toponymie est sujette à controverse, l'albanais et le serbe étant les deux langues officielles. Les modifications de valeur de tag sont en bleu.

Version Tag key	28	29	30	31	32	33
name	Malishevë - Mališevo	Malishevë	Malishevë	Malishevë/ Mališevo	Malishevë	Malishevë/ Mališevo
name:sq	Malishevë	Malishevë	Malishevë	Malishevë	Malishevë	Malishevë
name:sr-Latn		Mališevo	Mališevo	Mališevo	Mališevo	Mališevo

Ce travail exploratoire a été effectué manuellement, en parcourant sur la page des contributeurs bannis les messages émis par les modérateurs dans lesquels on retrouve parfois l'identifiant de la donnée ou du groupe de modifications (changeset) qui a mené au bannissement du contributeur, mais cette information n'est pas systématiquement renseignée. L'historique des données a

été obtenue avec l'application OSM Deep History<sup>8</sup>. Bien qu'il soit envisageable de fouiller systématiquement ces données, cela demande un travail conséquent d'analyse automatique de texte qui, selon nous, n'était pas primordial pour étudier les cas de vandalisme OSM. L'exploration des contributeurs bannis d'OSM montre que le bannissement ne concerne pas seulement les contributeurs ayant commis des actes de vandalisme sur les données OSM, mais également ceux qui ne respectent pas les règles de bonne conduite sur le projet : ne pas répondre aux commentaires adressés par d'autres contributeurs, importer massivement des données présentant une licence incompatible, ou participer à des guerres d'éditions. Finalement, le vandalisme sur OSM, selon sa définition traditionnelle, correspond à des cas très marginaux par rapport à l'ensemble des contributeurs bannis d'OSM. Les cas de vandalisme relevés à partir de la page des contributeurs bloqués correspondent aux cas les plus sérieux, néanmoins OSM encourage la communauté locale à résoudre les cas moins « sérieux » de vandalisme<sup>9</sup>.

### 3. Les approches de détection du vandalisme dans les bases de connaissance

Actuellement, l'outil de détection du vandalisme cartographique le plus abouti sur OSM est un système de décision multicritère (Neis *et al.*, 2012). Les règles de décision de cet outil prennent en compte la réputation du contributeur et l'historique des données OSM. Par conséquent, le système détecte facilement les objets créés par de nouveaux utilisateurs OSM, sans qu'ils soient forcément des cas avérés de vandalisme. Par ailleurs, d'après leurs résultats expérimentaux, le système a tendance à retourner un grand nombre de faux positifs, dont des éditions produites par des contributeurs expérimentés. Cela ne remet pas en cause la méthode de décision multicritère pour détecter le vandalisme, mais il faudrait trouver un ensemble de règles permettant de réduire le nombre de faux positifs dans ce système.

La détection de vandalisme dans les projets de crowdsourcing – tels que Wikipédia et Wikidata – est un sujet largement étudié dans la littérature (Adler *et al.*, 2011 ; Heindorf *et al.*, 2015). Les systèmes de détection proposés utilisent tous des méthodes de classification supervisée. Or, il n'existe pas de corpus de vandalisme pour les données spatiales, c'est pourquoi jusqu'à présent, l'approche supervisée n'a pas encore été envisagée pour détecter le vandalisme dans l'information géographique volontaire. Cette section explore les différentes approches possibles pour détecter le vandalisme dans les bases de connaissance. Tout d'abord, dans la section 3.1 nous analysons les travaux menés sur la détection du vandalisme dans les bases de connaissances (type Wikidata) par

8. <https://osmlab.github.io/osm-deep-history/>

9. <https://wiki.openstreetmap.org/wiki/FR:Vandalisme>

classification supervisée. Puis nous décrivons dans la section 3.2 notre démarche expérimentale pour détecter le vandalisme par une approche non-supervisée de détection d'anomalie.

### 3.1. Analyse des corpus de vandalisme

La construction d'un corpus de vandalisme consiste à annoter des données – par exemple **VRAI** pour une donnée vandalisée, **FAUX** sinon – à partir desquelles un système pourra apprendre les cas de vandalisme (jeu de données d'entraînement) ou évaluer le système de détection (jeu de données test). Alors que (Potthast, 2010) construit un corpus de vandalisme Wikipédia par annotation humaine via le service Amazon Mechanical Turk, des corpus récents de vandalisme de données Wikidata ont été construits par annotation automatisée.

Dans (Heindorf *et al.*, 2015; Sarabadani *et al.*, 2017), la première étape consistait à filtrer les révisions automatiques Wikidata, afin de ne conserver que les révisions manuelles. Parmi 24 millions d'éditions manuelles, Heindorf *et al.* (2015) ont considéré comme données vandalisées les contributions ayant subi une opération de révocation (*rollback*). La révocation est en réalité une opération autorisée pour un nombre restreint de contributeurs Wikidata, appelés les révocateurs<sup>10</sup>, consistant à annuler toutes les versions consécutives d'une donnée dans le but de la restaurer à un état précédant son état vandalisé. Sarabadani *et al.* (2017) suit une stratégie différente pour annoter automatiquement les données labélisées. En partant du principe que les données révoquées ne sont pas les seules données vandalisées dans un jeu Wikidata, l'obtention des données vandalisées dans ce corpus résulte d'un filtrage successif où les révisions faites par des contributeurs fiables ne sont pas considérées. En parcourant uniquement les éditions faites par des contributeurs non-fiables, les données issues d'une opération d'annulation – révocation ou autre opération de restauration autorisée à tous les contributeurs – sont labélisées **VRAI**.

On note que chaque corpus présente une part de faux positifs (révisions non vandalisées annotées **VRAI**) et de faux négatifs (révisions vandalisées annotées **FAUX**). Alors que le corpus proposé par (Heindorf *et al.*, 2015) a tendance à laisser passer de nombreux cas de vandalisme (32% de faux négatifs), celui de (Sarabadani *et al.*, 2017) détecte un grand nombre de faux cas de vandalisme (32% de faux positifs) en minimisant le nombre de faux négatifs (moins de 1%). Notons ici que la détection du vandalisme dans (Sarabadani *et al.*, 2017) a pour objectif d'aider les modérateurs à corriger les données vandalisées. Le système d'apprentissage fonctionne ici comme un outil d'aide à la décision, c'est-à-dire qu'en termes de performances, on exige en priorité un fort rappel, à la différence de la détection automatique du vandalisme, où l'on exige à la fois un fort rappel et forte précision. Tout compte fait, la tolérance aux faux

10. <https://www.wikidata.org/wiki/Wikidata:Rollbackers/fr>

positifs dans le corpus de vandalisme dépend de l'objectif final de la détection de vandalisme. Actuellement, il n'existe pas de corpus de vandalisme pour les données OSM, par conséquent, nous expérimentons la détection du vandalisme OSM par une méthode non-supervisée et dans le cadre d'une application d'aide à la décision (Table 3).

TABLE 3. *Différentes approches de détection du vandalisme*

Approche	Non supervisée	Supervisée
Application		
Aide à la décision	Notre expérience	(Sarabadani <i>et al.</i> , 2017)
Détection automatique du vandalisme		(Potthast, 2010)

### 3.2. Expérimentation préliminaire

Malgré la rareté des cas de vandalisme sur OpenStreetMap, le risque de provoquer le vandalisme est, en revanche, loin d'être nul puisque le mode de fonctionnement d'OSM offre la liberté au contributeur d'éditer la base de données géographiques comme bon lui semble. Toute nouvelle édition est directement chargée dans la base de données sans aucun contrôle préalable. Par conséquent, il est important de pouvoir détecter les cas de vandalisme au moins pour aider les contributeurs de bonne foi qui cherchent à réparer les données vandalisées.

Puisque les cas de vandalisme sur OSM sont très rares, qu'ils ne sont pas tous détectés et ne sont pas localisables, on tente de lancer une détection d'anomalie par une méthode de clustering sur les données OSM, où les anomalies doivent être des cas de vandalisme OSM à détecter. En effet, comme expliqué dans la partie 2.2., si les contributions vandalisées sont très marginales par rapport à l'ensemble des contributions des utilisateurs bannis, elles le sont encore plus par rapport à l'intégralité de la base cartographique. En outre, elles se distinguent des autres contributions car leur mauvaise qualité les rendent nuisibles à l'utilisation (navigation, analyse de biodiversité, analyse de morphologie urbaine, etc.). Par conséquent, une contribution vandalisée peut être considérée comme une anomalie à détecter dans un jeu de données OSM. Notre expérience vise à détecter automatiquement le vandalisme causé aux bâtiments OSM dans la ville d'Aubervilliers, en banlieue parisienne. Le jeu de données ne contient a priori pas de cas de vandalisme, par conséquent on ajoute manuellement des objets vandalisés (Figures 3 et 4) : 10 bâtiments sont tracés de manière artistique au milieu d'un cours d'eau et du cimetière d'Aubervilliers et 17 bâtiments fictifs de différentes tailles sont ajoutés dans des zones dépourvues de bâtiments (le polygone jaune de la Figure 4 indique que l'espace est actuellement une zone de construction).

Avant de lancer un algorithme de clustering sur le jeu de données ainsi vandalisé, on calcule un certain nombre de descripteurs sur les bâtiments OSM.

FIGURE 3. *Vandalisme synthétique de type artistique*FIGURE 4. *Vandalisme synthétique de type fantaisiste*

Comme le vandalisme synthétique consistait à ajouter des objets fictifs et artistiques, on se contente pour l'instant de calculer des variables géométriques (Table 4). Ainsi, l'algorithme de clustering regroupera les objets dont les attributs géométriques sont similaires en mettant de côté les bâtiments présentant des valeurs particulières, comme ceux de la Figure 3.

La détection d'anomalie utilise l'algorithme de clustering DENCLUE (implémentation issue de la bibliothèque Java SMILE), ce dernier étant invariant au bruit (Hinneburg, Gabriel, 2007). En fixant un paramètre de lissage gaussien  $\sigma$ , qui décrit l'influence d'une donnée dans son espace de données, et un

TABLE 4. Liste des variables géométriques calculées sur les bâtiments OSM  
 N.B. : PPRE signifie Plus Petit Rectangle Englobant.

Nom de la variable	Formule
$perimeter\_out\_of\_max$	$\frac{perimetre(bati)}{\max_{dataset}(perimetre)}$
$area\_out\_of\_max$	$\frac{aire(bati)}{\max_{dataset}(aire)}$
$shortest\_length\_out\_of\_perimeter$	$\frac{length(shortest\_edge_{bati})}{perimetre(bati)}$
$median\_length\_out\_of\_perimeter$	$\frac{length(median\_edge_{bati})}{perimetre(bati)}$
$elongation$	$\frac{largeur(PPRE_{bati})}{longueur(PPRE_{bati})}$
$convexity$	$\frac{aire(bati)}{aire(PPRE_{bati})}$
$compacity$	Miller's index: $\frac{4*\pi*aire(bati)}{perimetre(bati)^2}$

paramètre  $m$  correspondant au niveau de bruit, l'algorithme commence par construire un modèle de clustering à partir des variables d'entrée, puis prédit pour chaque bâtiment une classe selon le modèle construit. Un bâtiment particulier peut être classé dans un cluster à un élément ou directement comme une anomalie (*outlier*) si ses descripteurs sont complètement incohérents avec le modèle de clustering.

 TABLE 5. Résultats expérimentaux ( $\sigma = 0.005, m = 10250$ )

	Anomalies	Outlier	Partition à un élément
Vandalisme			
Artistique (total : 10)		1	9
Fictif (total : 17)		0	15
Non-vandalisme (total : 10315)		115	6080

On observe que 25 cas de vandalisme synthétique – artistique ou simplement fictif – sur les 27 ont été détectés : cela représente 92% des cas de vandalisme synthétique, ce qui constitue un résultat plutôt remarquable. En revanche, 60% des bâtiments a priori normaux ont été détectés comme des anomalies (*outliers* ou partitionnés dans un cluster de taille 1).

#### 4. Discussion

Dans le cas d'un outil d'aide à la décision, le système de détection du vandalisme doit minimiser en priorité le nombre de faux négatifs i.e. les données vandalisées qui ne sont pas labélisées comme telles. En cela, nos premiers résultats sont plutôt satisfaisants puisque notre expérience détecte moins de 8% de faux négatifs. Parmi ces faux négatifs, un minuscule bâtiment fictif n'a pas été reconnu comme une anomalie car le jeu de données contenait initialement un

grand nombre de petits bâtiments. En effet, les données OSM françaises sont en général issues de l'import massif de données du cadastre. Par conséquent, de nombreux éléments OSM cartographiés comme bâtiments sont en réalité des petits morceaux de bâtiments de formes singulières. Afin de distinguer les petits morceaux du cadastre de cas réels de vandalisme, il faudrait, dans l'idéal, faire en sorte d'utiliser des attributs géométriques qui ne fassent pas ressortir les spécificités géométriques des objets cartographiés. Ici, l'attribut *area\_out\_of\_max* n'était peut-être pas pertinent pour détecter les anomalies. De la même manière, il aurait peut-être fallu considérer une variable de densité de polygone comptant le nombre de sommets dans chaque bâtiment pour détecter tous les vandalismes artistiques.

Les variables d'entrée du clustering ne contenaient aucun attribut topologique qui aurait pu indiquer la présence de bâtiments vandalisés tracés dans un cours d'eau, dans une zone de construction et sur un cimetière. L'ajout de variables topologiques permettrait de détecter les bâtiments situés à des endroits inhabituels. D'autres descripteurs de type historiques, sémantiques et contributeurs auraient pu être ajoutés en entrée de l'algorithme, mais ceux-ci n'ont pas été implémentés. Dans un premier temps, nous cherchions à voir si les erreurs géométriques, dont nos actes de vandalisme, pouvaient être détectées comme des anomalies à partir de descripteurs purement géométriques.

Le cas d'étude présenté peut sembler simpliste puisque les cas de vandalisme synthétiques à détecter sont uniquement de type artistique et fantaisiste. Nous n'avons pas inséré dans le jeu de données OSM tous les types de vandalisme mentionnés dans la section 2.1 car l'objectif principal n'était pas de parvenir immédiatement à détecter tous les cas de vandalisme existants, mais plutôt d'évaluer la capacité d'un algorithme de détection d'anomalie par clustering à retrouver les cas de vandalisme artificiels, aussi simples soient-ils. Enfin, on ne s'attendait pas à détecter absolument tous les cas de vandalisme uniquement grâce à un partitionnement sur des variables géométriques. Ces premiers résultats sont donc plutôt encourageants pour la suite de nos recherches.

Par ailleurs, pour que le système soit véritablement une aide à la prise de décision, une réduction du nombre de faux positifs s'impose. Dans notre situation un faux positif correspond à un bruit que l'on cherche à éliminer. Pour cela, on peut filtrer les bâtiments normaux qui risquent d'être détectés par le système, en pré-qualifiant les objets présentant des spécificités géométriques, comme par exemple le centre commercial d'Aubervilliers, remarquable par sa grande taille ; ou des immeubles géométriquement remarquables par leur architecture originale. Cela nécessite une connaissance a priori du jeu de données initial. Cette étape de filtrage des faux positifs peut également être faite en sortie de l'algorithme de clustering : en prenant l'ensemble des données détectées comme anomalies, on pourrait filtrer les faux positifs par une seconde détection d'anomalie en prenant d'autres descripteurs en entrée. Dès lors que le nombre de faux positifs aura diminué, notre système de détection de vandalisme par clas-

sification non-supervisée pourra éventuellement s'étendre à une application de détection entièrement automatisée, mais cela ne pourra être envisagé que si l'on arrive à réduire drastiquement le nombre de faux positifs.

L'utilisation d'un algorithme d'apprentissage non-supervisé pour détecter avec précision le vandalisme semble être *a priori* une piste bien optimiste, surtout avec une définition où l'évaluation de l'intentionnalité, du dommage causé et du contexte, permet de le distinguer de l'erreur involontaire. Notre approche cherche à valider ou non l'hypothèse selon laquelle les contributions vandalisées présentent des caractéristiques aberrantes dans l'espace des données. En d'autres termes, on émet l'hypothèse que les cas de vandalisme sont des anomalies à détecter. La vérification de cette hypothèse s'effectue donc en choisissant les descripteurs les plus représentatifs de l'ensemble du jeu de données, de manière à ce que les contributions anormales ressortent à travers les valeurs de ces descripteurs. L'aspect intentionnel d'une erreur de saisie pourrait se modéliser par des descripteurs portant sur les contributeurs, mais ce modèle doit être vérifié expérimentalement.

La détection du vandalisme des données OSM par une classification supervisée nécessite de construire un corpus de contributions annotées. Dans le cadre d'une aide à la décision, on peut reprendre la piste proposée par (Sarabadani *et al.*, 2017) qui consiste à filtrer les contributions Wikidata faites par des contributeurs de confiance. Ainsi, les contributions douteuses de notre jeu de données seront récupérées pour être labélisées comme étant du vandalisme. Or, contrairement, à Wikidata, il n'y a pas de privilège accordée aux contributeurs OSM au niveau des opérations permises pour l'édition de données. Par conséquent, il sera nécessaire de mettre au point une méthode de qualification des contributeurs pour permettre d'évaluer leur niveau de fiabilité (Kessler, Groot, 2013 ; Truong *et al.*, 2018). Ce corpus de données annotées pourra contenir des cas de vandalisme synthétique similaires à ce que nous avons produits manuellement (Figures 3 et 4) mais également envisager de provoquer d'autres types de vandalisme synthétique. Par exemple, on pourrait cartographier des bâtiments qui – bien que fictifs – respecteraient les lois géographiques, c'est-à-dire qu'ils seraient géométriquement cohérents entre eux et avec les autres objets de la carte (Goodchild, Li, 2012). Ce genre de vandalisme fictif mais réaliste pourrait être détecté en changeant d'échelle d'étude, et ainsi, au lieu d'analyser les objets géographiques individuellement, il s'agirait d'analyser la cohérence entre des groupes d'objets pour identifier des cartographies imaginaires de villes ou de quartiers (Landrieu, Simonovsky, 2017).

Enfin, l'éventualité de passer à une détection totalement automatique du vandalisme OSM par classification supervisée est encore loin. En effet, celle-ci suppose de construire un corpus solide de vandalisme OSM qui ne laisse pas de place à l'approximation : toutes les contributions OSM du corpus devront être annotées de manière à ce que le nombre de faux positifs et de faux négatifs soit nul. Dans ce cadre-là, l'annotation automatique proposée par Sarabadani



*et al.* n'est pas assez satisfaisante pour construire un tel corpus. Il faudra alors chercher un moyen de faire mieux que les méthodes proposées jusqu'à présent dans la littérature.

## 5. Conclusion et perspectives

Dans cet article, notre travail s'est attaché à la définition même de ce qu'est le vandalisme et en quoi l'évaluation de ses trois composantes – dommage causé, intentionnalité et contexte – devient un réel défi pour développer sa détection de manière automatique. L'état de l'art, plutôt riche quant à la détection de vandalisme des bases de connaissances participatives par approche supervisée, a permis de révéler un manque de recherche sur la détection par classification non-supervisée. Les premières expériences de détection par classification non-supervisée ont permis de rendre compte de l'importance du choix des descripteurs en entrée de l'algorithme de clustering. En effet, ces attributs ne peuvent pas être de simples indicateurs de qualité mais doivent être spécialement conçus pour révéler la particularité des cas de vandalisme. Nos résultats expérimentaux nous ont encouragés à réfléchir à des moyens de réduire le nombre de faux positifs et de faux négatifs. En améliorant nos résultats expérimentaux, nous pourrions alors envisager d'étendre la détection de vandalisme pour répondre à des objectifs plus ambitieux tels que la construction d'un corpus de vandalisme voire une détection "tout automatique" du vandalisme.

Dans nos travaux futurs, il s'agira d'améliorer les résultats de notre expérience initiale, en faisant varier dans un premier temps l'influence des paramètres  $\sigma$  et  $m$  de l'algorithme DENCLUE. Dans un second temps, on pourra tester d'autres algorithmes de clustering, tels que DBScan ou BIRCH, pour vérifier si ceux-ci sont plus efficaces pour détecter le vandalisme. De plus, pour sortir du contexte français des données OSM où les bâtiments sont cartographiés par morceaux, on prévoit de lancer la même expérience sur un jeu de données provenant d'un autre pays tel que l'Allemagne où les bâtiments sont généralement cartographiés manuellement. Dans ce nouveau jeu de données, les cas de vandalisme qui n'étaient pas détectés parmi les bâtiments d'Aubervilliers pourraient l'être parmi ceux d'une ville allemande. Comme il a été mentionné précédemment, la détection d'autres types de vandalisme sera aussi étudiée, telle que le vandalisme par édition de tag ou par suppression d'objet. Dans ce cas, il sera nécessaire d'enrichir les descripteurs de données OSM par d'autres types d'attributs (historique, topologique, contributeur, sémantique) et lorsque ce sera possible, des indicateurs d'appariement avec des données de référence. A nouveau, il faudra choisir judicieusement ces descripteurs, en prenant soin d'en limiter le nombre afin de ne pas se trouver piégé par la malédiction de la grande dimension. Enfin, à partir du moment où les résultats expérimentaux présenteront beaucoup moins de faux positifs, la construction d'un corpus de données de vandalisme des données OSM pourra être envisagée, en poursuivant la réflexion sur les différents points abordés dans la discussion de cet article. Ce

corpus pourra alors être utilisé pour détecter le vandalisme des données OSM par une approche supervisée.

## Bibliographie

- Adler, Alfaro L., Mola-Velasco S. M., Rosso P., West A. G. (2011). Wikipedia vandalism detection: Combining natural language, metadata, and reputation features. In A. Gelbukh (Ed.), *Computational linguistics and intelligent text processing*, vol. 6609, p. 277–288. Berlin, Heidelberg, Springer Berlin Heidelberg. Consulté sur [http://dx.doi.org/10.1007/978-3-642-19437-5\\_23](http://dx.doi.org/10.1007/978-3-642-19437-5_23)
- Ballatore A. (2014, 1<sup>er</sup> août). Defacing the map: Cartographic vandalism in the digital commons. *The Cartographic Journal*, vol. 51, n° 3, p. 214–224. Consulté sur <http://dx.doi.org/10.1179/1743277414y.00000000085>
- Goodchild M. F., Li L. (2012, 28 mars). Assuring the quality of volunteered geographic information. *Spatial Statistics*.
- Heindorf S., Potthast M., Stein B., Engels G. (2015). Towards vandalism detection in knowledge bases. In *Proceedings of the 38th international acm sigir conference on research and development in information retrieval - sigir '15*, p. 831–834. ACM Press. Consulté sur <http://dx.doi.org/10.1145/2766462.2767804>
- Hinneburg A., Gabriel H. H. (2007). DENCLUE 2.0: Fast clustering based on kernel density estimation. In M. R. Berthold, J. S. Taylor, N. Lavrac, M. R. Berthold, J. S. Taylor, N. Lavrac (Eds.), *Ida*, vol. 4723, p. 70–80. Springer. Consulté sur [http://dx.doi.org/10.1007/978-3-540-74825-0\\_7](http://dx.doi.org/10.1007/978-3-540-74825-0_7)
- Ivanovic S. (2018). *Une approche basée sur la qualité pour mettre à jour les bases de données géographiques de référence à partir de traces GPS issues de la foule*. Thèse de doctorat non publiée, Université Paris Est.
- Keßler C., Groot R. de. (2013). Trust as a proxy measure for the quality of volunteered geographic information in the case of OpenStreetMap. In D. Vandenbroucke, B. Bucher, J. Cromptvoets (Eds.), *Geographic information science at the heart of europe*, p. 21–37. Springer International Publishing. Consulté sur [http://dx.doi.org/10.1007/978-3-319-00615-4\\_2](http://dx.doi.org/10.1007/978-3-319-00615-4_2)
- Landrieu L., Simonovsky M. (2017). *Large-scale point cloud semantic segmentation with superpoint graphs* (vol. abs/1711.09869). Published 2017 in ArXiv.
- Neis P., Goetz M., Zipf A. (2012, 22 novembre). Towards automatic vandalism detection in OpenStreetMap. *ISPRS International Journal of Geo-Information*, vol. 1, n° 3, p. 315–332. Consulté sur <http://dx.doi.org/10.3390/ijgi1030315>
- Noucher M. (2014, septembre). *L'information géographique volontaire : vers un renouvellement de la fabrique cartographique institutionnelle ?*
- Potthast M. (2010). Crowdsourcing a wikipedia vandalism corpus. In *Proceedings of the 33rd international acm sigir conference on research and development in information retrieval*, p. 789–790. New York, NY, USA, ACM. Consulté sur <http://dx.doi.org/10.1145/1835449.1835617>
- Sarabadani A., Halfaker A., Taraborelli D. (2017). Building automated vandalism detection tools for wikidata. In *Proceedings of the 26th international conference*

*on world wide web companion - www '17 companion*, p. 1647–1654. ACM Press.  
Consulté sur <http://dx.doi.org/10.1145/3041021.3053366>

Touya G., Brando C. (2013). *Detecting Level-of-Detail inconsistencies in volunteered geographic information data sets* (vol. 48) n° 2.

Truong Q.-T., Touya G., Runz C. de. (2018). Building social networks in volunteered geographic information communities: What contributor behaviours reveal about crowdsourced data quality. In P. Fogliaroni, A. Ballatore, E. Clementini (Eds.), *Proceedings of workshops and posters at the 13th international conference on spatial information theory (cosit 2017)*, p. 125–131. Springer International Publishing. Consulté sur [http://dx.doi.org/10.1007/978-3-319-63946-8\\_25](http://dx.doi.org/10.1007/978-3-319-63946-8_25)

Zimbardo P. G. (1971). *A Social-Psychological analysis of vandalism: Making sense of senseless violence*. Rapport technique. Stanford University, Department of Psychology.

---

# Résolution des conflits lors de la conception collaborative de cubes OLAP pour des observatoires citoyens

Amir Sakka<sup>2</sup>, Sandro Bimonte<sup>1</sup>, Lucile Sautot<sup>4</sup>, Guy Camilleri<sup>2</sup>,  
Pascale Zaraté<sup>2</sup>, Aurelien Besnard<sup>3</sup>

1. UR TSCF, IRSTEA de Clermont-Ferrand, 9 Avenue Blaise Pascal, 63170 Aubière, France, [sandro.bimonte@irstea.fr](mailto:sandro.bimonte@irstea.fr)

2. IRIT, Université Paul Sabatier, Toulouse, France

3. LPO Aquitaine, 433 Chemin de Leyssotte, 33140 Villenave-d'Ornon, France

4. UMR TETIS, AgroParisTech, CIRAD, CNRS, IRSTEA 500 rue Breton, 34093 Montpellier Cedex 5, France, [lucile.sautot@agroparistech.fr](mailto:lucile.sautot@agroparistech.fr)

---

**RÉSUMÉ.** DANS LE CONTEXTE DE L'INFORMATION GÉOGRAPHIQUE VOLONTAIRE (VGI), LES VOLONTAIRES NE SONT PAS IMPLIQUÉS DANS LES PROCESSUS DÉCISIONNELS. DE PLUS, LES SYSTÈMES VGI N'OFFRENT PAS D'OUTILS PUISSANTS POUR MENER DES ANALYSES TEMPORELLES. C'EST POURQUOI, DANS CET ARTICLE, NOUS PROPOSONS D'UTILISER LES SYSTÈMES D'INFORMATION DÉCISIONNELS POUR ANALYSER LES DONNÉES VGI, ET NOUS PROPOSONS LA DÉFINITION D'UNE NOUVELLE MÉTHODOLOGIE DE CONCEPTION DES ENTREPÔTS DE DONNÉES, QUI PERMET L'IMPLICATION DES VOLONTAIRES DANS LA DÉFINITION DES BESOINS ANALYTIQUES SUR LES DONNÉES VGI. NOS PROPOSITIONS ONT ÉTÉ TESTÉES SUR UN CAS D'ÉTUDE RÉEL CONCERNANT LA BIODIVERSITÉ.

**MOTS-CLÉS :** OLAP, Entrepôt de données, Information géographique volontaire (VGI), Système de prise de décision en groupe

---

**ABSTRACT.** IN THE CONTEXT OF VOLUNTEERED GEOGRAPHIC INFORMATION (VGI), VOLUNTEERS ARE NOT INVOLVED IN DECISION-MAKING PROCESSES. IN ADDITION, VGI SYSTEMS DO NOT OFFER POWERFUL TOOLS FOR TEMPORAL ANALYSES. THEREFORE, IN THIS ARTICLE WE PROPOSE TO USE BUSINESS INTELLIGENCE SYSTEMS TO ANALYZE VGI DATA, AND WE DEFINE A NEW DATA WAREHOUSE DESIGN METHODOLOGY, WHICH ALLOWS THE INVOLVEMENT OF VOLUNTEERS IN THE DEFINITIONS OF ANALYTICAL NEEDS ON VGI DATA. OUR PROPOSALS HAVE BEEN TESTED ON A REAL CASE STUDY ON BIODIVERSITY.

**KEYWORDS:** OLAP, Data Warehouse, Voluntary Geographic Information, Groupe Decision Making System

---

## 1. Introduction

Le VGI (Information géographique volontaire) est la mobilisation d'outils pour créer, assembler, et disséminer les données géographiques fournies par des

volontaires [8]. Classiquement, les volontaires sont des producteurs de données actifs, et des consommateurs passifs des analyses de données fournies par les organismes ou les entreprises concernées. Ce paradigme représente une barrière importante pour le développement d'observatoires basés sur des contributions volontaires, car les producteurs de données peuvent se sentir exclus du processus de prise de décision [5]. De plus, comme souligné dans [1], la VGI ne propose pas de fonctionnalités analytiques pour de grands volumes de données spatiales. En revanche, les systèmes OLAP Spatial (SOLAP) sont pertinents pour les analyses basées sur l'exploration de jeux de données géographiques massifs stockés dans un entrepôt de données spatiales [7, 4].

Plus le modèle de l'entrepôt de données reflète les besoins des décideurs, plus ces décideurs utilisent les données [4,6]. Cet article a donc pour objectif de "transformer" les volontaires producteurs de données en analystes de données. Plusieurs méthodologies de conception d'entrepôts de données ont été proposées dans la littérature [2,6]. Néanmoins, quand les décideurs sont des volontaires, ils peuvent avoir des parcours et des fonctions variés (scientifiques, citoyens, etc.), ce qui peut conduire à de multiples interprétations contradictoires des mêmes besoins. Quand les décideurs ont des buts différents, il devient difficile de maintenir un consensus entre eux, comme le montrent les travaux en ingénierie des besoins [3,11]. Ainsi, les besoins définis par ces "décideurs volontaires" présentent *des définitions d'éléments multidimensionnels qui sont soit erronées, soit non pertinentes* [11]. De fait, impliquer ces utilisateurs particuliers dans les méthodologies de conception d'entrepôt de données existantes n'est pas possible car ces méthodologies nécessitent une connaissance avancée des principaux concepts de la modélisation multidimensionnelle, et font l'hypothèse que les utilisateurs sont effectivement impliqués dans l'ensemble du projet.

Nous proposons donc une méthodologie innovante de conception collaborative d'entrepôt de données, utilisant un système d'aide à la décision en groupe (GDSS), qui aide les volontaires à choisir les besoins les plus pertinents parmi les besoins exprimés. En effet, les outils de GDSS sont conçus pour aider un groupe engagé dans un processus de décision collectif et collaboratif. Ce type de système n'a pas encore été utilisé pour la conception d'entrepôt de données [9].

## 2. Cas d'étude

Dans le cadre du projet ANR VGI4Bio, nous avons mobilisé deux bases de données VGI (Visionature et l'Observatoire Agricole de la Biodiversité - OAB) afin de construire une application SOLAP, qui permettra l'analyse d'indicateurs de biodiversité dans les milieux agricoles. Visionature et l'OAB impliquent respectivement environ 7700 et 1500 volontaires, qui produisent les données. A cette étape du projet, nous avons identifié 12 volontaires pouvant se prêter à une expérimentation de la méthodologie proposée.

La Figure 1 montre un modèle multidimensionnel défini par un volontaire concernant l'analyse de l'abondance de plusieurs espèces animales. Ce modèle permet de formuler des requêtes SOLAP comme : “*Quel est l'abondance totale des oiseaux selon l'altitude, l'espèce et la semaine ?*”.

L'objectif de notre proposition étant la résolution de conflits au sein des besoins définis par des volontaires, nous proposons d'illustrer la notion de conflit à partir de ce cas d'étude. Par exemple, pour la mesure de l'abondance de certaines espèces (“abondance”) comme les papillons, le protocole d'acquisition prévoit que l'opérateur observe les animaux en se déplaçant le long d'une ligne, d'une distance donnée, ou pendant une durée définie (environ 100 mètres pour les papillons). Ainsi, la mesure de l'abondance des papillons n'a de sens pour l'analyse biologique, que si elle est accompagnée de la durée ou de la longueur de l'observation.

Cet exemple nous montre que certains éléments multidimensionnels, considérés comme nécessaires par les volontaires lors de la définition du modèle, ne sont pas cohérents vis à vis du domaine d'application. Ces conflits ne sont pas causés par les données sources, mais sont le résultat de disparités de connaissance et d'expertise dans le domaine d'application du volontaire. Ainsi, ces conflits ne peuvent pas être résolus par un outil automatique, mais seulement par des experts du domaines.

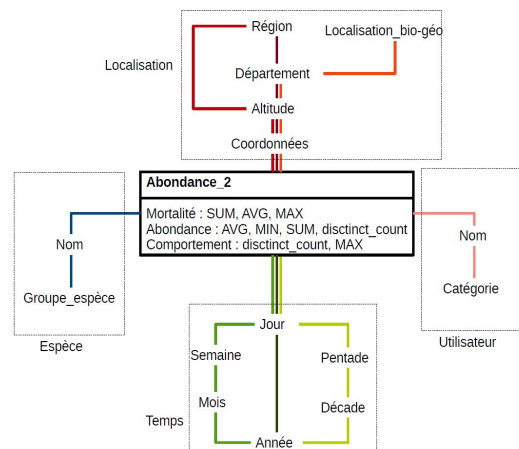


Figure 1. Exemple de modèle multidimensionnel conçu par un volontaire.

### 3. Méthodologie proposée

L'objectif de méthodologie est de résoudre les conflits engendrés par la définition des modèles multidimensionnels par les volontaires : “*Les éléments multidimensionnels (mesures, dimensions, hiérarchies) proposés par les volontaires sont-ils tous nécessaires ?*”.

Comme nous l'avons illustré avec le cas d'étude, les conflits engendrés par la définition des besoins par les volontaires doivent être tranchés au regard du domaine

d'application. De plus, comme indiqué précédemment, les volontaires participent ponctuellement à la collecte des données et ne sont pas impliqués en permanence dans le projet VGI. Pour pallier à ces limites dans notre approche, les modèles multidimensionnels sont proposés à un ensemble de volontaires particuliers appelés "committers", qui sont complètement impliqués dans le projet. Les committers décident si les besoins recueillis auprès des volontaires (sous forme de modèles multidimensionnels) seront implémentés ou non, en jugeant de la pertinence de ces besoins, grâce à leur expertise dans le domaine d'application.

Dans cette section, nous présentons la méthode générale de résolution des conflits. Cette méthode présente les différentes étapes que doivent suivre les committers pour statuer sur un modèle proposé.

Les committers définissent leur niveau d'expertise pour le cube selon leurs compétences dans le domaine d'application. Concrètement, les committers s'évaluent eux-mêmes. Ce niveau d'expertise permet de rendre prioritaires les choix des committers avec les compétences les plus appropriées vis-à-vis de l'évaluation du cube. Les committers évaluent successivement les indicateurs, les dimensions, les hiérarchies de chaque dimension, les niveaux de chaque hiérarchie. Chaque élément qui n'est pas jugé assez utile par le groupe est éliminé du modèle. Notons qu'un élément (cube, dimension, hiérarchie) dont les sous-éléments (indicateurs, hiérarchie, niveau) ont tous été éliminés, est lui même supprimé du modèle. Après cela, les committers doivent évaluer l'utilisabilité du cube avec chaque dimension retenue [10], car le nombre de dimensions utilisées affecte l'utilisabilité du cube, et le processus de prise de décision. Dans cette optique, on ajoute, en commençant par les dimensions les mieux notées par les committers, les dimensions au cube consécutivement. Après chaque ajout de dimension, le cube résultant est présenté aux committers. Par ce moyen, les committers peuvent explorer le cube avec la dimension ajoutée, décider de l'utilisabilité du cube avec la nouvelle dimension, et choisir de la conserver ou pas. Enfin, les committers votent l'implémentation du cube résultant.

#### 4. Expérimentation

Notre approche a été implémentée dans une architecture ROLAP, utilisant Postgresql comme base de données, Mondrian comme serveur OLAP et JRubik comme client OLAP. Les méthodes de l'algorithme ont été implémentés dans le GDSS GRUS [9]. GRUS est un GDSS web, qui permet d'organiser le processus de choix des éléments à conserver ou à retirer des modèles. Il permet aux committers de voter pour les différents éléments de manière indépendante.

Dans cette section nous présentons les résultats d'une première expérimentation menée sur notre cas d'étude. Pour cette expérimentation, qui aura vocation à être complétée dans des travaux futurs, les committers sont trois experts appartenant à la LPO, la DREAL et AgroParisTech. Ils ont des profils différents (écologie,

gestionnaire environnementale et ingénieur en agronomie) et ils connaissent très bien les données sources.

Pour cette première expérience, notre méthodologie a été appliquée au cube de Figure 2a, qui est un cube qui a été conçu par un volontaire. Le cube résultat est montré en Figure 2b.

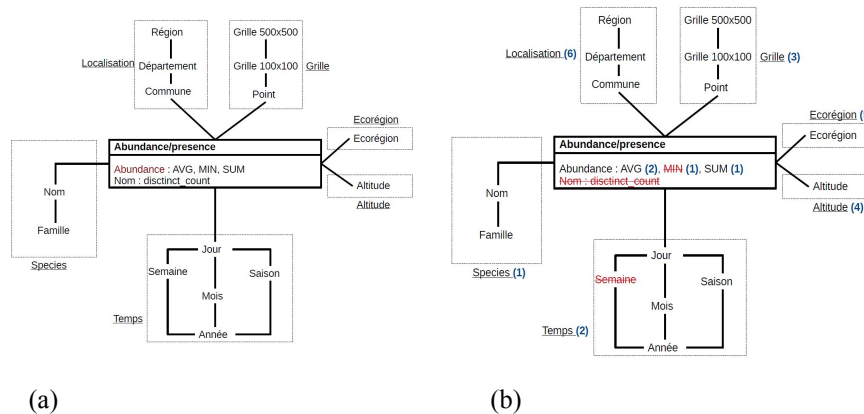


FIGURE 2. *Modèle multidimensionnel pour la biodiversité défini par un volontaire*

Les indicateurs “Abondance+Min” et “nom+Distinct Count” et le niveau “Semaine” ont été éliminés du cube de la Figure 2b. Les chiffres notés à côté des noms des dimensions indiquent l’ordre d’évaluation des dimensions. Les committers ont ensuite voté le cube de la Figure 2b en décidant de l’implémenter.

Le processus d’évaluation a duré 2 heures, et a été réalisé par visioconférence. Nous avons ensuite questionné les committers sur la méthodologie. Ils estiment utile l’organisation et donc l’évaluation des composantes du cube comme nous l’avons définie. Ils ont aussi apprécié le système collaboratif fourni par GRUS. Par contre, ils ont fait émerger la nécessité de pouvoir apporter des modifications au cube proposé au début de l’évaluation, ce qui représente une future amélioration de notre algorithme.

## 5. Conclusions et travaux futurs

Dans cet article, nous proposons une nouvelle méthode de résolution de conflits, adaptée à la conception collaborative d’entrepôt de données. Cette méthode est une étape importante pour l’implication des volontaires dans la définition des besoins analytiques concernant des données VGI. La méthode proposée permet à des volontaires sans aucune compétence en système OLAP de participer au processus de conception de ce type d’outil. Cette méthode a été implémentée, ce qui a permis de mener les premières expérimentations. Ces premières expérimentations, bien que nécessitant d’être complétées, ont permis de valider la méthode sur un cas d’étude réel concernant la biodiversité dans les milieux agricoles. Ainsi, à l’avenir, nous proposerons une extension des critères utilisés par notre approche collaborative en se



basant sur des métriques qualitatives définies pour la satisfaction des utilisateurs d'entrepôts de données.

**Remerciements.** Ces travaux ont reçus le soutien financier du projet ANR-17-CE04-0012 VGI4Bio. Nous remercions tous les volontaires qui participent au projet VGI4Bio.

## Références

1. Bimonte, S., Boucelma, O., Machabert, O., Sellami, S.: A new Spatial OLAP approach for the analysis of Volunteered Geographic Information. *Computers, Environment and Urban Systems*. 48, 111–123 (2014).
2. Cravero, A., Sepúlveda, S.: Multidimensional Design Paradigms for Data Warehouses: A Systematic Mapping Study. *Journal of Software Engineering and Applications*. 07, 53 (2013).
3. Egyed, A., Grunbacher, P.: Identifying requirements conflicts and cooperation: how quality attributes and automated traceability can help. *IEEE Software*. 21, 50–58 (2004).
4. Kimball, R., Ross, M.: *The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence Remastered Collection*. John Wiley & Sons (2016)
5. Levrel, H., Fontaine, B., Henry, P.-Y., Jiguet, F., Julliard, R., Kerbirou, C., Couvet, D.: Balancing state and volunteer investment in biodiversity monitoring for the implementation of CBD indicators: A French example. *Ecological Economics*. 69, 1580–1586 (2010).
6. Romero, O., Abelló, A.: A Survey of Multidimensional Modeling Methodologies. *IJDWM*. 5, 1–23 (2009).
7. Stefanovic, N., Han, J., Koperski, K.: Object-based selective materialization for efficient implementation of spatial data cubes. *IEEE Transactions on Knowledge and Data Engineering*. 12, 938–958 (2000).
8. Sui, D.Z., Elwood, S., Goodchild, M. eds: *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*. Springer Netherlands (2013)
9. Zaraté, P.: *Tools for Collaborative Decision-Making: Zaraté/Tools for Collaborative Decision-Making*. John Wiley & Sons, Inc. (2013)
10. Golfarelli, M., Rizzi, S.: Data warehouse testing: A prototype-based methodology. *Information & Software Technology* 53(11): 1183–1198 (2011)
11. Pohl, K.: *Requirements Engineering: Fundamentals, Principles, and Techniques*. Springer Publishing Company, Incorporated (2010)

---

# FabSpace 2.0: une innovation pédagogique qui met les géodonnées au coeur des apprentissages

Josiane Mothe<sup>1</sup>

ESPE, IRIT, URM5505 CNRS, 118 route de Narbonne, 31062 Toulouse, France

Josiane.Mothe@irit.fr

---

**RÉSUMÉ.** *FabSpace 2.0 est un Fablab qui utilise les données du spatial. L'idée est de faire travailler ensemble, autour d'un objet et d'un projet commun, des personnes avec des compétences variées. FabSpace 2.0 est centré sur l'innovation axée sur les géo-données, avec une attention particulière aux données d'observation de la Terre captées par le programme Européen Copernicus. Ainsi, à l'Université les étudiants et les chercheurs de toutes disciplines sont sollicités pour contribuer à résoudre des problèmes en lien avec des sujets de société posés par des structures extérieures (publiques, privées ou associatives). Ils bénéficient de ressources pédagogiques de différents formats, sur des sujets techniques mais aussi une sensibilisation à l'entrepreneuriat. L'idée est d'amener l'innovation au cœur de l'université et de rendre les étudiants les acteurs de cette innovation en proposant de nouveaux services ou applications basées sur les géo-données. Ainsi les étudiants apprennent de façon innovante.*

**ABSTRACT.** *FabSpace 2.0 is FabLab-like place that uses space data. We make people with varied skills from various domains working together on a common application idea. FabSpace 2.0 focuses on geo-data driven innovation, with a particular attention to Earth observation data from the European Copernicus program. At the University students and researchers from all disciplines are asked to contribute to solve societal problems posed by external structures (public, private, or associative). They benefit from learning resources of different formats, on technical subjects but also an awareness to entrepreneurship. The idea is to bring innovation to the heart of the university and to make students the actors of this innovation by proposing new services or applications based on geo-data. Students learn in a different way.*

**MOTS-CLÉS :** *Géo-données, Innovation pédagogique, Observation de la Terre, Application innovantes, Copernicus*

**KEYWORDS:** *Geo-Data, pedagogical innovation, Earth observation, innovative application, Copernicus*

---

## 1. Introduction: une innovation pédagogique basée sur les géo-données

Cet article court présente FabSpace 2.0, une innovation pédagogique qui est développée autour des géo-données dans différentes universités (Del Frate, Carbone *et al.*, 2017).

FabSpace 2.0 est la combinaison des différentes pratiques pédagogiques actuelles. La démarche fait largement appel aux outils numériques. Cela recouvre par exemple des dispositifs d'enseignement en ligne (vidéos, tutoriels, MOOCs, plateforme Moodle) qui permettent de dispenser des connaissances de façon personnalisée, mais aussi des enseignements en présentiel dispensés par des professionnels du secteur du spatial (Aerospace Valley, CNES, Thalès Alenia Space, CLS, ATOS...) ou des chercheurs, ainsi que des événements de créativité et d'idéation tels que des mapathons et des hackathons durant lesquels se constituent des équipes projets multidisciplinaires.

Ainsi, grâce à la dématérialisation des supports d'enseignement, le contenu en ligne est accessible de partout et à toute heure, et chaque étudiant peut composer son programme en fonction de son parcours, de ses connaissances précédentes et des besoins d'apprentissage liés à la problématique choisie. Les échanges en direct avec des professionnels extérieurs et avec des chercheurs favorisent l'émergence d'idées d'applications spatiales et la participation à des concours permet aux étudiants de se mettre dans la peau d'un porteur de projet, jusqu'à l'imagination du plan d'affaires et de ce que pourrait être sa future entreprise.

L'apprentissage est au service de la problématique à résoudre. L'étudiant est motivé par l'enjeu, l'implication étant le meilleur facteur de réussite (Eseryel *et al.*, 2014). La validation des Unités d'Enseignement auxquelles la démarche FabSpace 2.0 est associée au résultat auquel l'étudiant contribue, plutôt qu'à sa propre réussite académique.

Le public actuel est constitué d'étudiants de niveau Master 1 et Master 2, de doctorants et de post-doctorants dans toutes les disciplines et plus particulièrement dans des domaines applicatifs s'appuyant sur des géo-données (environnement, géomatique, aménagement du territoire, etc.).

Il s'agit d'augmenter la capacité du public ciblé à utiliser les services de FabSpace afin de disséminer parmi eux la volonté de développer de nouvelles applications à hauts enjeux sociétaux et économiques à partir du traitement des données. Les ressources existantes permettent de cibler des groupes d'étudiants et de chercheurs susceptibles d'apporter une forte valeur ajoutée aux activités d'innovation au sein des services FabSpace.

## 2. Dispositifs mis en œuvre

Nous pouvons citer cinq enjeux principaux que nous développons dans cette section: (1) Intégration des services de FabSpace 2.0 dans les programmes et activités existants à l'université (2) Initiation aux géo-données et à leurs applications (3) Promotion de l'esprit d'entreprise chez les étudiants (4) Développement des connaissances des chercheurs en leadership de l'innovation et (5) Accès par les étudiants et chercheurs à des contenus de formation spécifiques sur l'entrepreneuriat et les questions techniques.

**Intégration des services de FabSpace 2.0 dans les programmes et activités existants des universités.** Nous intervenons dans les modules projets en partenariat avec les formations de niveau Master en mathématique informatique, en sciences géomatiques. Nous faisons travailler les étudiants en autonomie dans des modules projets en leur apportant des formations complémentaires qu'ils peuvent trouver dans FabSpace 2.0 (partenariat avec les entreprises Airbus, ATOS, CLS, C-S Communication Systems, FeelObject 3D, IdGEO, Oracle, Orange, Snap Planet, TerraNIS, Thalès Alenia Space).

**Initiation aux géo-données et à leurs applications.** Nous organisons des conférences et des ateliers autour de sujets d'actualité. Des experts du domaine sont sollicités pour présenter aux participants les enjeux et les méthodes associées. Par exemple, nous avons abordé les sujets de la déforestation, du 7<sup>ème</sup> continent ou du suivi des espèces marines lors de conférences. Ces conférences ont lieu entre midi et deux, tous les jeudis de sorte que cette activité puisse s'intégrer facilement dans les emplois du temps des étudiants.

**Promotion de l'esprit d'entreprise chez les étudiants.** Nous sensibilisons les étudiants pour qu'ils deviennent acteurs de l'entrepreneuriat, de l'innovation, de la recherche et du développement. En effet, dans la plupart des formations, les étudiants n'ont pas de formation à l'entrepreneuriat. Des solutions existent déjà en dehors de FabSpace 2.0 pour accompagner les jeunes entrepreneurs dans leur démarche dans les universités comme des dispositifs jeunes pousses. A l'Université Toulouse III – Paul Sabatier il s'agit des dispositifs Catalyseur et Pépite. Lorsque l'idée est plus mature, les étudiants s'adressent aux dispositifs de création de start-ups. A Toulouse, le pôle de compétitivité Aerospace Valley en tant qu'incubateur d'entreprises du domaine spatial, véritable tremplin vers une phase d'incubation, est une possibilité.

Ces structures et dispositifs n'ont cependant pas vocation à être embarquées dans les formations. De plus, ils s'adressent essentiellement aux étudiants qui se considèrent déjà comme entrepreneurs. La démarche FabSpace 2.0, elle, s'intéresse spécifiquement aux étudiants qui sont dans une phase très amont de leur projet, pour leur donner le goût du monde entrepreneurial. Au sein du FabSpace de Toulouse, cela prend la forme d'un programme de sensibilisation à l'entrepreneuriat (Startech), d'appui à la gestion de projet grâce à la disponibilité du FabSpace manager et de toute l'équipe, d'un programme de maturation du

projet entrepreneurial réservé aux idées d'applications spatiales les plus avancées (Bootcamp) mais également de la mise en place de Hackathons qui ont pour objet soit de développer des applications qui répondent à une demande d'un partenaire (par exemple une ONG), soit de proposer de nouvelles idées dans un domaine d'application (agriculture, énergie, etc.).

**Développement des connaissances des chercheurs en leadership de l'innovation.** Les chercheurs formés ont de meilleures connaissances en matière de leadership de l'innovation et sont invités à devenir des utilisateurs de FabSpace 2.0 pour compléter leur programme de formation par une activité d'apprentissage par la pratique. Un programme de formation est proposé dans l'offre de programmes de formation thématiques pour l'apprentissage tout au long de la vie des chercheurs.

**Accès par les étudiants et chercheurs à des contenus de formation spécifiques sur l'entrepreneuriat et des questions techniques.** Différentes applications sont en cours de développement au sein de groupes projets multi-disciplinaires et multi-compétences, impliquant des étudiants, des chercheurs et des entreprises.

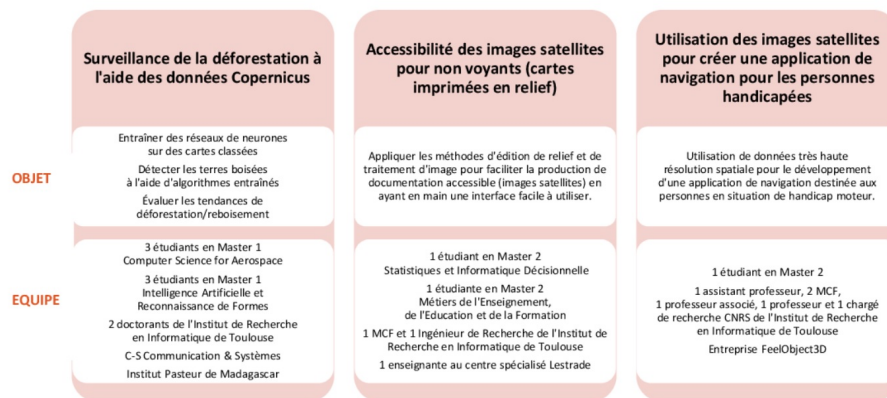


FIGURE 1. Exemples de projets développés dans le cadre de l'innovation pédagogique FabSpace.

### 3. Partenariat et reproductibilité

Dans cette innovation pédagogique, les partenaires locaux jouent un rôle clé. A Toulouse, nous pouvons citer:

- Le pôle de compétitivité Aerospace Valley, qui compte 812 membres dont 513 PME, co-organise de très nombreuses animations et fait appel à des in-

tervenants industriels clefs du domaine du spatial pour partager leurs connaissances (CNES, Airbus, ATOS, Thalès Alenia Space). Aerospace Valley pilote par ailleurs l'ESA BIC Sud France, incubateur de l'ESA ;

- La PME TerraNIS fournit un outil de visualisation des données, forme à son utilisation et conseille les étudiants sur les ressources d'intérêt ;

- L'organisme de formation IdGEO, spécialisé en géomatique, propose des apports théoriques spécifiques pour permettre la montée en compétence des étudiants, et leur donne un éclairage pratique sur les aspects applicatifs des données d'observation de la Terre ;

- Le service RUS (Research and User Support for Copernicus Core Product) de la commission européenne est au service des utilisateurs de Sentinel. Il est géré par l'ESA et opéré par C-S. Il donne accès à un traitement en ligne des données et à d'autres solutions techniques ainsi qu'à un support technique et à des sessions de formation.

La démarche FabSpace 2.0 est actuellement opérationnelle dans six universités européennes (Del Frate, Mothe *et al.*, 2017) (France, Italie, Belgique, Allemagne, Pologne, Grèce) et est en train d'être déployée dans 14 nouveaux FabSpaces dans toute l'Europe (France, Italie, Espagne, Pays-Bas, Lituanie, République Tchèque, Chypre) et même au-delà (Arménie, Cameroun). En France, trois nouveaux établissements d'enseignement supérieur ont été sélectionnés pour le déployer : l'ENSEIRB-Matmeca à Bordeaux, l'IMT Atlantique à Brest et l'Université Nice Côte d'Azur. En 2019, elle va s'étendre au niveau mondial. La transférabilité de la démarche FabSpace 2.0 est ainsi complètement opérationnelle grâce aux nombreux retours d'expériences cumulés depuis 2016 et à la mutualisation de toutes les ressources qui s'y prêtent (documentation technique, supports de communication, jeux de données, etc.). Un ouvrage est en préparation qui permettra à de nouveaux FabSpaces de se développer.

La démarche est également associée à des événements majeurs du monde du spatial comme le concours international ActInSpace dont FabSpace 2.0 est un partenaire majeur par la proposition de défis et la préparation des éléments techniques d'intérêts pour leur résolution (jeux de données, espaces de travail virtuels, plate-forme de visualisation). Nous pouvons également citer le challenge CLEF pour lequel la population d'une région devait être estimée (Arenas *et al.*, 2017).

L'actualité du projet est accessible sur le site internet [www.fabspace.eu](http://www.fabspace.eu) et sur les réseaux sociaux (Twitter, LinkedIn, facebook). Au niveau local, les actualités de la démarche FabSpace 2.0 font l'objet de nombreuses communications via la presse (communiqués et articles) et via des lettres d'information électroniques, ainsi que par le site internet [www.irit.fr/FabSpace/fr](http://www.irit.fr/FabSpace/fr) et les réseaux sociaux (Twitter, facebook).

#### 4. Conclusion

FabSpace développe une nouvelle forme de dispositif d'apprentissage en mettant l'application et l'innovation au coeur du dispositif (Marantos *et al.*, 2017). L'utilisation des géo-données s'est avérée motivante pour les étudiants car ils ont ainsi accès à de nouvelles formes de données concrètes et qui ouvrent des champs d'application énormes.

Le public actuel est constitué principalement d'étudiants de niveau Licence 3, Master 1 et Master 2, de doctorants et de post-doctorants dans des disciplines variées; certains dans des domaines applicatifs tels que l'environnement, la géomatique, l'aménagement du territoire, etc. Nous visons à étendre l'impact de cette innovation pédagogique en proposant des activités dans d'autres types de formation comme en architecture par exemple.

Nous avons réalisé une enquête de satisfaction auprès de nos utilisateurs, pour laquelle l'anonymat des répondants était garanti. Il ressort de cette enquête que la moitié seulement des utilisateurs possèdent des connaissances d'informatique ou de programmation; que les utilisateurs préfèrent les sujets autour de la protection de l'environnement, le changement climatique, les villes intelligentes ou l'agriculture, mais sont moins intéressés par des sujets comme les assurances ou l'énergie par exemple. FabSpace 2.0 a reçu le soutien de l'Europe dans le cadre du programme Horizon 2020 (programme Recherche et Développement Horizon 2020 - convention n°693210).

#### Bibliographie

- Arenas H., Islam M. B., Mothe J. (2017). Overview of the ImageCLEF 2017 Population Estimation (Remote) Task. In *International Conference of the CLEF Association*, vol. 1866. CEUR Workshop Proceedings. Consulté sur [http://ceur-ws.org/Vol-1866/invited\\_paper\\_2.pdf](http://ceur-ws.org/Vol-1866/invited_paper_2.pdf)
- Del Frate F., Carbone F., Mothe J., Baker A., Paraskevas I., Soundris D. *et al.* (2017). The FABSPACE 2.0 Project For Geodata-Driven Innovation. In *ESA Earth Observation Open Science*. ESA. Consulté sur <https://livestream.com/ESA/OpenScience2017>
- Del Frate F., Mothe J., Barbier C., Becker M., Olszewski R., Soundris D. (2017). FabSpace 2.0: The Open-Innovation Network for Geodata-driven Innovation. In *International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE.
- Eseryel D., Law V., Ifenthaler D., Ge X., Miller R. (2014). An investigation of the interrelationships between motivation, engagement, and complex problem solving in game-based learning. *Journal of Educational technology & society*, vol. 17, n° 1.
- Marantos C., Paraskevas I., Siozios K., Mothe J., Menou C., Soundris D. (2017, mai). FabSpace 2.0: A Platform for Application and Service Development based on Earth Observation Data (regular paper). In *International Conference on Modern Circuits and Systems Technologies (MOCAST)*. IEEE. Consulté sur <http://doi.org/10.1109/MOCAST.2017.7937657>

---

# Analyser la structure de dispositifs numériques de mise à disposition de données géographiques sur le Web

Johan Desbonnet<sup>1</sup>, Françoise Gourmelon<sup>1</sup>, Christophe Claramunt<sup>2</sup>

1. LETG-Brest (UMR 6554 CNRS), Université Bretagne Occidentale  
IUEM, Rue Dumont d'Urville, 29280 Plouzané cedex  
[johan.desbonnet@univ-brest.fr](mailto:johan.desbonnet@univ-brest.fr) ; [francoise.gourmelon@univ-brest.fr](mailto:francoise.gourmelon@univ-brest.fr)

2. Institut de Recherche de l'Ecole Navale (IRENaV)  
BCRM Brest, CC 600, Ecole Navale, 29240 Brest, cedex 9  
[christophe.claramunt@ecole-navale.fr](mailto:christophe.claramunt@ecole-navale.fr)

---

**RESUME.** Dans le cadre du suivi et de la gestion du trait de côte, le besoin d'informations et de connaissances accessibles sur le Web est actuellement décrit comme un enjeu par les pouvoirs publics français. Néanmoins sur cette thématique, la stratégie des promoteurs des Infrastructures de Données Géographiques (IDG) et des plateformes de données ouvertes de même que leurs contenus restent encore peu connus. Cet article décrit une approche méthodologique pour explorer les volets structurels de l'accès à l'information géographique numérique à partir de l'analyse des liens hypertextuels et des services de consultation de dispositifs numériques susceptibles d'être mobilisés dans le suivi et la gestion du trait de côte. L'analyse des composants structurels des dispositifs sur la toile tels que le volet hypertextuel et les services Web révèle un cheminement parfois complexe et peu diversifié pour accéder à une information géographique pertinente sur cette thématique.

**ABSTRACT.** In the context of shoreline management, inferring information and knowledge from the Web are considered by French public authorities as a crucial issue. However, Spatial Data Infrastructures promoters' strategies and content are still unknown. The preliminary research presented in this paper provides a methodological approach that explores the availability of geographical information as exhibited by the hypertext and Web services structure of geographical digital infrastructures. Analysis of hypertext and web services shows sometimes complex path to access the geographical information on digital platforms studied.

**MOTS-CLES :** Information géographique numérique, Infrastructures de Données Géographiques (IDG), plateformes de données ouvertes, structure hypertextuelle, services Web, mise en réseau, gestion et suivi du trait de côte

**KEYWORDS:** Digital geographical information, Spatial Data Infrastructures (SDI), Open Data platforms, hypertext structure, Web services, network, shoreline management

---



## 1. Introduction

Face aux enjeux présents dans les espaces bordant le trait de côte (risques, érosion de la biodiversité, aménagement du territoire), la mise en place et le renouvellement de projets en faveur de la production et de l'accès à une information géographique numérique (Litto 3D, indicateur national d'érosion, ortholittorale, observatoires, etc.) sur des dispositifs en ligne sont érigés en axes prioritaires par les pouvoirs publics depuis le début des années 2000 (Bersani et al., 2006 ; Georis-Creuseveau, 2014). Ces dispositifs, qui regroupent dans notre étude les Infrastructures de Données Géographiques (IDG)<sup>1</sup> et les plateformes de données ouvertes<sup>2</sup> ont permis la mise en ligne de l'information géographique sur le Web. Ils sont le résultat d'un investissement progressif des pouvoirs publics, des sociétés privées et des organisations non gouvernementales dans le domaine de l'accès à l'information à l'échelle mondiale, européenne (convention d'Aarhus, directive Inspire) et nationale (directives Cada et PSI, Loi Notre, Loi pour une république numérique dite aussi loi Lemaire).

Ces dispositifs, de plus en plus nombreux sur le Web 2.0, ont pris au cours du temps, de leurs évolutions techniques, et des objectifs fixés par leurs promoteurs (services de l'État, établissements publics) des dénominations variées : bases de données géographiques, API cartographiques<sup>3</sup>, Géoportails... Ils sont souvent décrits comme des technologies visant à faciliter l'accès et les usages de l'information dans un cadre d'action publique (suivi et gestion des zones côtières, urbanisme, risques, etc.) et peuvent être développés pour et par des acteurs spécifiques (Pornon, 2015 ; Georis-Creuseveau et al., 2016). Les travaux portant sur ces infrastructures de données géographiques numériques ont montré que de tels dispositifs pouvaient être influencés par différents facteurs techniques (normes, standards), politiques, organisationnels, économiques et sociétaux (Rajabifard, 2002 ; Noucher, 2013 ; Georis-Creuseveau, 2014). L'Association Française pour l'Information Géographique (AFIGEO) recensait 65 dispositifs (IDG) en 2014<sup>4</sup> tandis que le programme de recherche Geobs<sup>5</sup>, qui s'intéresse aux dynamiques de circulation de

<sup>1</sup> Les Infrastructures de Données Géographiques, régies par des politiques et des normes, ont l'objectif de favoriser la mise en réseau des producteurs et des utilisateurs de l'information géographique (Rajabifard, 2002 ; Georis-Creuseveau, 2014 ; Noucher et al., 2017).

<sup>2</sup> Les plateformes de données ouvertes étudiées sont le résultat d'un investissement des gouvernements et agences gouvernementales dans l'ouverture de leurs données publiques. Ce sont des dispositifs qui s'appuient le plus souvent sur des algorithmes, des méthodes au service d'une politique et qui mettent à disposition ou renvoient vers des données de producteurs sous une licence libre pour permettre le développement d'autres applications (Colin et Verdier, 2015).

<sup>3</sup> Une API est une interface de programmation pouvant être considérée comme une application technique support de fonctionnalités spécifiques (cartographiques, statistiques et graphiques) et qui prend le plus souvent la forme de service web (Mericksay, 2011).

<sup>4</sup><http://www.afigeo.asso.fr/documentation/publications.html?download=718:catalogue-des-idg-03-juin-2014>

<sup>5</sup> <http://geobs.cnrs.fr/>, <http://www.geobs.cnrs.fr/manifesto.pdf>

l'information géographique au sein de ces dispositifs, en recensait 45 aux échelons national et régional (Noucher et al., 2017). En parallèle et en lien avec la mouvance « open data », des plateformes de données ouvertes telles que data.gouv et geodatagouv<sup>6</sup> ont été mises en place par le gouvernement français. Enfin, dans le cadre spécifique français de la stratégie nationale de gestion intégrée du trait de côte, un certain nombre d'infrastructures de données géographiques qualifiées « d'observatoires et de systèmes d'information sur le littoral (SILs) » (Suanez et al., 2012), ont été identifiées par les pouvoirs publics (en gras et soulignés dans le Tableau 1).

Ces différentes définitions et recensements sur les IDG et plateformes de données ouvertes françaises nous ont conduits à sélectionner un corpus de 43 dispositifs dont certains sont déployés spécifiquement dans le cadre de la thématique du suivi et de la gestion du trait de côte (Tableau 1).

TABLEAU 1. Corpus étudié (n=43)

Type	Echelle	Nom
IDG identifiées comme « observatoires et systèmes d'information sur le littoral » (n= 15)	Supranationale (n=3)	<b><u>Emodnet</u> ; <u>Sedatanet</u> ; <u>API « Coastline » (absence de site Web dédié)</u></b>
	Nationale (n=10)	<b><u>API « Rivages »</u> ; <u>Anemoc</u> ; <u>Candhis</u> ; <u>Bosco</u> ; <u>Geolittoral</u> ; <u>Onml</u> ; <u>Datashom</u> ; <u>Sextant</u> ; <u>Cartomer</u> ; <u>Dynalit</u></b>
	Régionale/ Interrégionale (n=2)	<b><u>Observatoire du Littoral Bas-Normand (Oliban)</u> ; <u>Réseau d'Observation du Littoral Normand et Picard (Rolnp)</u></b>
Autres IDG généralistes et thématiques françaises (n = 27)	Supranationale (n=2)	Geoportail Inspire Europe ; Portail de données de l'Agence Européenne de l'Environnement
	Nationale (n=17)	API Vigicrues ; Base de données Historiques sur les Inondations (BDHI) ; Georisques ; Infoterre ; EauFrance ; Inventaire National du Patrimoine Naturel (INPN) ; Géoportail national de l'urbanisme <sup>o</sup> ; Carmen Nature France ; ADéLie ; Cartélie, Geo-IDE ; Geocatalogue national ; Géoportail national ; Observatoire des territoires ; Portail de données publiques de Météofrance ; Applications Geoclip géo (hors ligne depuis début 2018) et Geoclip stat du Service de l'Observation et des Statistiques
	Régionale/ Interrégionale (n=7)	Geobretagne ; Bretagne Environnement ; Kartenn Bretagne ; Geonormandie ; Geopal ; Indigeo ; Osuna

<sup>6</sup> <https://www.data.gouv.fr/fr/> <https://geo.data.gouv.fr/fr/>

	Communale/ Intercommunale (n=1)	Géocatalogue de Lorient Agglomération
Plateforme de données ouvertes (n =1)	Nationale (n=1)	Datagouv

S'intéresser aux questions de disponibilité de l'information géographique sur une thématique donnée et sur des dispositifs numériques tels que les IDG et les plateformes de données ouvertes permet de traiter des questions relatives à ses modes de communication dans « l'écosystème numérique » qu'est le Web (Cardon, 2011). Les aspects dits structurels de tels dispositifs numériques s'appuient sur les principes de fonctionnement de l'Internet et du Web 2.0 et constituent à ce titre des enjeux de recherche pour appréhender les processus qui orientent le parcours de l'information géographique numérique dans la gouvernance informationnelle.

Le Web peut être considéré comme un espace façonné par des intentionnalités et des actions d'acteurs qui affichent, diffusent et partagent de l'information dans un réseau interconnecté (Beaude, 2012 ; Graham et Dutton, 2014) et sur des thématiques fixées. À ce titre, Internet et la toile constituent un objet d'investigation (Gautreau et al., 2013 ; Desbois, 2015 ; Severo et Venturini, 2016 ; Noucher, 2017) permettant de caractériser certaines pratiques et stratégies autour de la mise à disposition, de la visibilité et de la mise en réseau de l'information numérique (Gautreau et al., 2013 ; Graham et Zook, 2013 ; Noucher et al., 2017).

Cette recherche vise donc à analyser la contribution d'un ensemble de dispositifs mettant à disposition de l'information géographique numérique sur le Web utilisable sur des enjeux de suivi et de gestion du trait de côte en France. Pour ce faire, nous proposons, à partir de méthodes de recherche dites numériques (Rogers, 2009, 2015 ; Plantin et Monnoyer-Smith, 2014) d'explorer des aspects renvoyant aux caractéristiques structurelles de ces dispositifs. Cette démarche est développée à partir d'une analyse de la structure hypertextuelle et des services Web disponibles (Rogers, 2015) que nous quantifions à partir d'indicateurs statistiques et de l'étude des liens entre ces dispositifs.

## **2. Le Web comme espace structurel d'investigation de dispositifs numériques en lien avec l'information géographique : éléments méthodologiques**

### ***2.1. L'analyse hypertextuelle : un indicateur des ressources offertes par les dispositifs numériques***

La structure hypertextuelle des sites Web des IDG et des plateformes de données ouvertes est une première porte d'entrée technique sur leurs contenus. Elle permet à

l'utilisateur de parcourir et d'accéder à des ressources informationnelles sur le site Web de l'IDG ou de la plateforme de données ouvertes qu'il consulte, à partir de liens structurés renvoyant vers différents espaces de consultation sous la forme de pages (par exemple, document sur un autre site Web, espace de téléchargement, cartes, données en ligne etc.). Une telle armature hypertextuelle offre la possibilité d'observer la trame globale du dispositif et non de se concentrer sur un seul composant que serait le visualiseur cartographique, le catalogue ou bien encore l'information géographique diffusée (carte, graphique, etc.). Les hyperliens sont considérés par certains courants de recherche sur le numérique (Digital Methods Initiative<sup>7</sup>) comme des indicateurs de richesse et de diversité informationnelle, notamment les liens sortants (Rogers, 2009). Ils mettent en évidence la redirection du site vers d'autres sources potentielles d'information auxquelles l'utilisateur n'aurait pas accès dans le dispositif visité. Les hyperliens, en permettant la mise en réseau des différents espaces informationnels, peuvent favoriser l'accès à « de nouveaux espaces de connaissances » (Pinède, 2014). Pour les chercheurs en informatique (Kleinberg, 1999 ; Page, Brin et al., 1999 ; Chakrabarti, 2002), la nature documentaire du Web peut être caractérisée par sa structure topologique dont les liens entrants et sortants sont révélateurs de phénomènes de centralité et de hiérarchie. Les liens entrants sont considérés comme des indicateurs « d'autorité » sur le Web c'est-à-dire que la page Web d'un dispositif possède une autorité d'autant plus forte que la somme des degrés sortants des pages qui la mentionne est élevée. Les liens sortants mettent en évidence le phénomène de « hubs » définis comme des pages Web contenant un grand nombre de liens sortants vers d'autres pages Web qui elles-mêmes possèdent une forte autorité. C'est pourquoi les liens hypertextes sont aussi définis dans la littérature comme le résultat de « politiques » (Rogers, 2009) et de stratégies (Gautreau et al., 2013 ; Pinède, 2014) qui peuvent caractériser l'association ou une distance entre les acteurs porteurs des dispositifs.

Afin d'étudier la structure hypertextuelle de notre corpus, la méthode appliquée consiste à explorer, indexer et catégoriser les pages Web, les liens renvoyés et les sites Web découverts à partir d'un programme similaire à ceux utilisés par les moteurs de recherche : un « crawler » ou moissonneur (exemple : le « Googlebot » du moteur de recherche Google). Un « crawler » est un logiciel qui permet de collecter les pages et les liens du site Web visité et ainsi de reconstituer « un espace de cheminement pour l'utilisateur du Web » (Rogers, 2009). Le « crawler » indexe les pages jusqu'à une certaine distance dans le site (ou niveau de profondeur), c'est-à-dire un nombre de clics en partant de la page du site Web que l'on renseigne au « crawler » (ici  $n = 2$ )<sup>8</sup>. Le « crawler » utilisé est le logiciel Hyphe<sup>9</sup>. Une fois le

<sup>7</sup> Le « Digital Methods initiative » ou « courtant de l'Internet » est un groupe de recherche européen qui s'intéresse aux stratégies de diffusion de l'information par les hyperliens, <https://www.digitalmethods.net/Digitalmethods/WebHome>

<sup>8</sup> Ce choix est motivé par le fait que la profondeur de 2 correspond à l'espace où l'on trouve les liens les plus importants (prioritaires) dans l'environnement des sites Web (Jacomy, Ghitalla, 2007). C'est généralement à ce niveau que les pages les plus importantes se situent, ce qui peut s'expliquer par le fait que plus les pages sont profondes moins les moissonneurs des moteurs de recherche leur accordent de l'importance.

crawl réalisé en partant de la page d'accueil de chaque dispositif (IDG ou plateforme de données ouvertes), on obtient des résumés sur le nombre de pages Web crawlées, le nombre total de pages Web et de sites Web découverts et le nombre de liens hypertextes présents sur les pages du dispositif visité. Par exemple en partant de la page d'accueil de l'IDG Geolittoral, le programme Hyphe a moissonné 434 pages Web internes à une distance de 2 clicks de profondeur. Il a découvert et indexé 14230 liens hypertextes, 13593 adresses de pages Web et 47 sites Web extérieurs. Ces valeurs extensives permettent d'analyser statistiquement les ressources hypertextuelles offertes par chaque dispositif. Afin de caractériser cette ressource, on propose d'étudier la dispersion de la structure hypertextuelle des dispositifs numériques (calcul des valeurs centrales et des paramètres de dispersion, forme de la distribution, représentation graphique).

## ***2.2. Analyse de graphe sur les liens hypertextes et sur les services de consultation des données (services Web, conventions) : un indicateur de la structure en réseau des dispositifs sur la toile***

Les notions de réseau renvoient à des questions d'interopérabilité pour les dispositifs numériques sur le Web (Gasser, 2015), notamment ceux dédiés à l'information géographique (Rajabifard, 2002 ; Noucher, 2013 ; Georis-Creuseveau, 2014). En France, la mise en réseau par l'interopérabilité des IDG institutionnelles est promue par la directive Inspire et par la politique de la donnée de l'État. La structure en réseau de ces dispositifs peut dès lors être caractérisée par une analyse et une comparaison des graphes au niveau hypertextuel et au niveau des services de consultation de la donnée (services Web, autres). Le graphe est la figure la plus employée dans la représentation de la topologie du Web. C'est avant tout un objet mathématique défini « comme un ensemble fini de sommets  $N$  et un ensemble fini de liens  $L$  pouvant être vide, noté  $G = (N ; L)$  » (Beaugitte, 2010).

Afin de représenter le réseau hypertextuel, un crawl a été lancé à partir du logiciel Hyphe sur les pages d'accueil des 43 dispositifs du corpus en août 2017. À partir des résultats des enquêtes du projet de recherche Geobs<sup>10</sup> sur les services Web de moissonnage (standard CSW) des IDG institutionnelles françaises et du recensement des autres possibilités de consultation (services OGC standardisés de visualisation WMS, WFS et de téléchargement WPS, conventions), on a cartographié l'ensemble des liens entre les dispositifs. Les relations hypertextuelles et de service ont été cartographiées à l'aide du logiciel Gephi 0.9.1. La spatialisation des dispositifs, c'est-à-dire la représentation de l'ensemble de leurs relations au niveau hypertextuel ( $n = 43$  dispositifs avec 114 liens) et au niveau des relations de visualisation, de catalogage et de téléchargement ( $n=40$ <sup>11</sup> avec 152 liens) vise à faciliter la lecture du positionnement des dispositifs sur le Web à un instant donné.

<sup>9</sup> Hyphe est développé par le Sciences Po Médialab dans le cadre du projet Hypertext corpus initiative en 2010 (Girard et al., 2012), <http://hyphe.medialab.sciences-po.fr/>.

<sup>10</sup> <http://geobs.cnrs.fr/>

<sup>11</sup>  $N = 40$  dispositifs dans le réseau de services et non 43, car les IDG thématiques et généralistes européennes n'ont pas été intégrées à ce niveau d'analyse.

Cette visualisation s'appuie sur un même algorithme de spatialisation basé sur les forces (Futcherman, Reingold, 1991) qui permet la comparaison à partir d'une représentation des nœuds selon des principes d'attraction (les plus liés s'attirent) et de répulsion (les moins liés sont repoussés à la périphérie). Enfin, on a appliqué des algorithmes employés dans la caractérisation de la nature topologique du Web et dans l'indexation de son contenu pour mesurer la centralité de chaque IDG et de la plateforme de données ouvertes (« Autorité » et « Hub » de Kleinberg (1999) et algorithme de classement « PageRank » de Page et al. (1999)).

### **3. Analyse de la structure hypertextuelle et de la structure en réseau des dispositifs numériques : résultats**

#### ***3.1. La distribution des pages Web, liens et entités Web : une première mesure de la ressource hypertextuelle***

Concernant notre corpus, les résumés statistiques des distributions hypertextuelles (liens, pages Web et entités<sup>12</sup> Web) montrent qu'une personne souhaitant consulter les dispositifs jusqu'à une distance de 2 clicks, a accès en moyenne à 22449 pages Web, 11791 liens hypertextuels et à 129 sites Web externes. La valeur modale dans les trois distributions est assez faible, c'est-à-dire que la valeur qui revient le plus fréquemment pour les dispositifs dans les trois séries statistiques est égale à 0 pour les liens découverts, égale à 2 pour le nombre de pages Web découvertes et égale à 0 pour le nombre d'entités Web découvertes. On trouve dans les trois cas des distributions dissymétriques à gauche (médiane très inférieure à la moyenne) c'est-à-dire qu'une majorité des dispositifs se situe dans les valeurs faibles. Ainsi comme on peut le voir sur les diagrammes en bâtons (Figure 1), certains dispositifs ne présentent aucune ou peu de pages (18 dispositifs renvoyant à un nombre de pages Web découvertes inférieur à 100), peu ou aucun lien (19 dispositifs avec un nombre de liens découverts inférieur à 50) et ne renvoient qu'à peu de sites Web différents (20 dispositifs renvoyant à un nombre de sites Web externes inférieur à 5). Seuls quelques dispositifs nationaux et régionaux détiennent des valeurs élevées en nombre de liens, de pages découvertes et d'entités Web (INPN, EauFrance, Datagouv, Geobretagne, Bretagne environnement).

Si l'on compare la répartition des dispositifs dans les trois distributions (pages Web, liens hypertextes, entités Web) selon un découpage en effectifs égaux (méthode des quantiles), on distingue un premier groupe de dispositifs caractérisé par un faible nombre de liens et de pages Web (dispositifs dont la valeur est inférieure à Q1 sur la figure 1). On retrouve dans ce premier groupe 5 dispositifs en lien avec le suivi et la gestion du trait de côte. Le second groupe (dispositifs avec une valeur supérieure à Q3) rassemble des dispositifs qui présentent un grand nombre de pages Web et de liens. Ce sont des IDG thématiques nationales (INPN, SIE, Vigicrues) et régionales (Bretagne environnement), une IDG généraliste régionale (Geobretagne) et la plateforme nationale de données ouvertes (Datagouv).

---

<sup>12</sup> Une entité web est un site Web.

Concernant les dispositifs identifiés dans la thématique du suivi et de la gestion du trait de côte, aucun dispositif ne se situe dans les 25 % des effectifs aux valeurs les plus élevées pour l'ensemble des indicateurs (nombre de liens, de pages Web et d'entités Web externes). Ce sont l'IDG européenne (Emodnet), les IDG nationales (Geolittoral et Sextant) et l'IDG interrégionale (ROLNP) qui présentent la structure hypertextuelle à la plus grande diversité en termes d'indicateurs hypertextuels (dans 2 indicateurs sur 3, les valeurs sont supérieures à Q1).

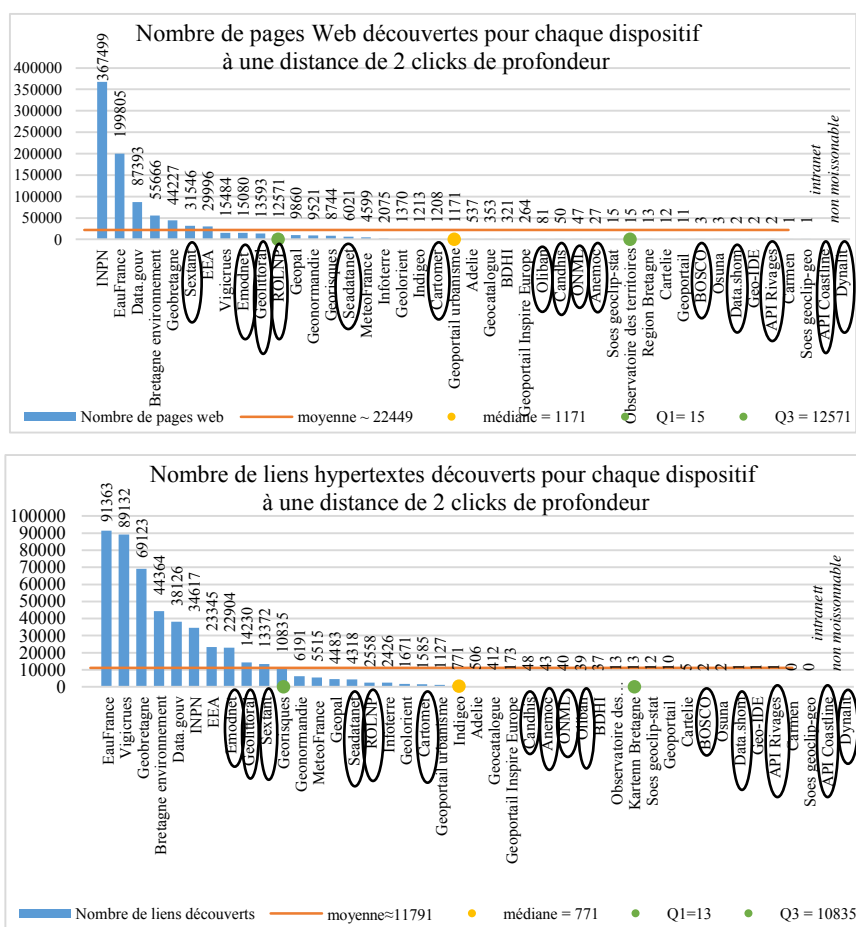


FIGURE 1. Deux des trois indicateurs de la structure hypertextuelle : nombre de pages Web (en haut) et nombre de liens (en bas) (les systèmes d'information sur le littoral – observatoires sont entourés en noir)

### **3.2. Analyse de la composante réseau des dispositifs numériques : un second indicateur pour caractériser l'aspect structurel**

Le second indicateur structurel met en regard le niveau hypertextuel avec le niveau de services de consultation de la donnée (services Web et conventions) par l'analyse de graphe des liens entre dispositifs numériques (Figure 2). Dans les deux niveaux de graphe, la densité<sup>13</sup> est relativement faible sur l'ensemble du réseau ( $< 0,1$ ), ce qui indique que les dispositifs du corpus n'exploitent pas le potentiel maximal de connexion. De plus, on trouve un nombre élevé de composantes<sup>14</sup> connexes c'est-à-dire de sous-graphes (25 au niveau hypertextuel et 23 au niveau des services) pouvant être identifiés comme des communautés ; et des composantes faiblement connexes (7 au niveau hypertextuel et 4 au niveau des services). Les dispositifs les plus centraux dans le graphe (Figure 2) sont des IDG nationales thématiques (INPN, EauFrance) ainsi que des infrastructures généralistes nationale (Géocatalogue) et régionale (Geonormandie) et la plateforme de données ouvertes (Data.gouv). Dans les distributions de degrés entrants et sortants, on observe qu'une majorité de dispositifs situés dans le premier quartile de la distribution (25% des effectifs aux plus forts degrés entrants et sortants) sont des infrastructures nationales généralistes et thématiques.

Ce sont les dispositifs nationaux Geolittoral et l'ONML identifiés par la stratégie nationale de gestion intégrée du trait de côte et les systèmes d'information mer et littoral (Sextant, Cartomer et Datashom) qui sont les plus centraux dans les deux graphes (Figure 2). D'un point de vue hypertextuel, si la majorité des dispositifs (thématiques, généralistes, territoriaux) sont connectés aux IDG identifiées par la stratégie nationale de gestion intégrée du trait de côte, cela ne saurait masquer une faible intensité des liens (31 liens sur les 114 liens recensés). Au niveau du graphe des services, d'une part les possibilités de consultation des données d'un dispositif du trait de côte à un autre dispositif restent limitées (52 liens sur 152 liens recensés). D'autre part, la majorité des relations par l'intermédiaire des services de consultation de la donnée sont le plus souvent cantonnées à une possibilité de consultation unique par la présence d'un seul type de service (téléchargement ou moissonnage ou visualisation). Enfin au niveau de l'analyse de graphe, 6 dispositifs sont en situation d'isolat<sup>15</sup> au niveau hypertextuel. Deux dispositifs identifiés dans la stratégie nationale de gestion intégrée du trait de côte (Bosco et Oliban) et 2 dispositifs en lien avec l'information géographique numérique mer et littoral (Candhis et Anemoc) sont isolés au niveau du réseau des services.

Si l'on compare les 10 scores les plus élevés des indicateurs de centralité (Autorités, « hubs », « PageRank »), ce sont les IDG nationales thématiques

<sup>13</sup> « La densité (*density*) d'un graphe désigne le rapport entre le nombre de liens présents et le nombre de liens possibles elle varie entre 0 (graphe vide, sans aucun lien) et 1 (graphe complet). » (Beaugitte, 2013)

<sup>14</sup> « Un composant correspond à un sous graphe non connecté aux autres sous-graphes. » (Beaugitte, 2010).

<sup>15</sup> Un nœud ou sommet est dit en situation d'isolat lorsqu'il n'est connecté à aucun autre nœud.



(EauFrance, INPN) et généralistes (Géoportail, Géo catalogue) qui se partagent les plus forts scores de centralité au niveau topologique. Parmi les dispositifs dédiés à l'information géographique numérique sur le trait de côte, on retrouve ainsi quatre IDG nationales (Geolittoral, Onml, Datashom et Sextant) dans les plus forts scores de centralité. Au niveau des indicateurs d'autorité et de « hubs », seul un dispositif (Datashom) possède un score élevé (autorité) dans les deux réseaux (hypertexte et service). Les 4 dispositifs Geolittoral, Onml, Datashom et Sextant possèdent un score élevé de « PageRank » dans les deux réseaux.

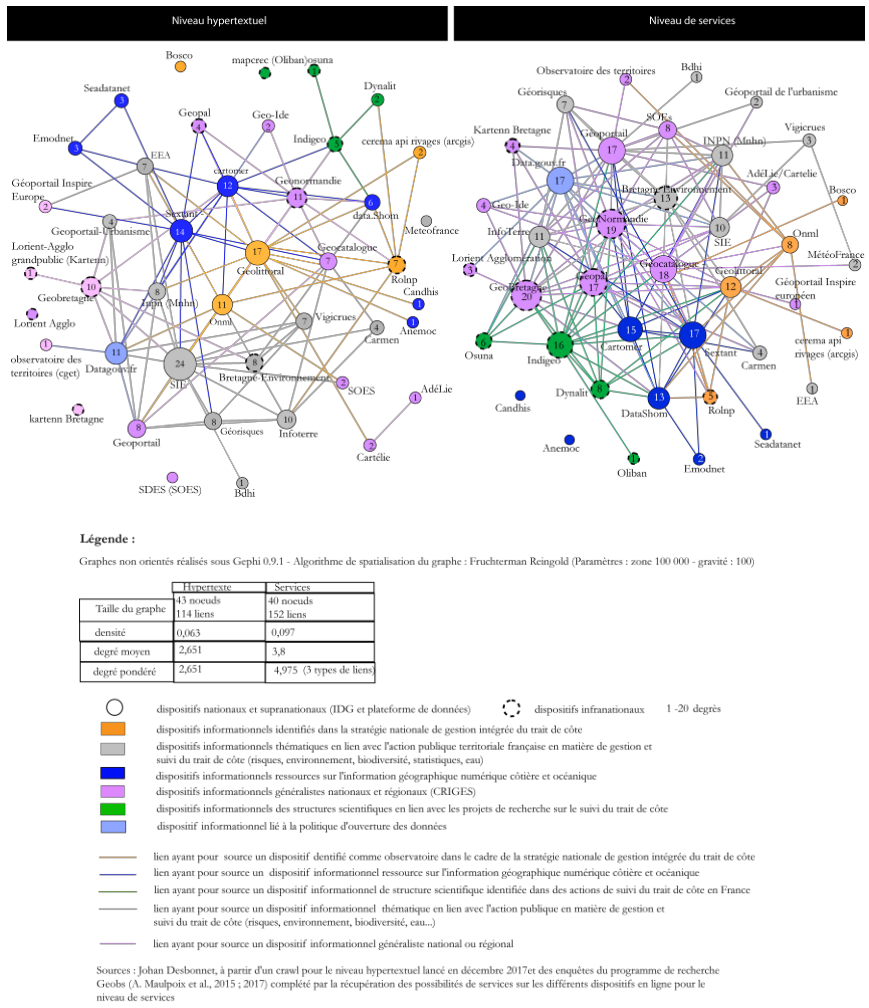


FIGURE 2. Mise en regard des graphes des liens entre dispositifs au niveau hypertextuel et au niveau des services de consultation de données

#### 4. Discussion

Nous proposons un protocole méthodologique visant à fournir des indicateurs en lien avec les caractéristiques structurelles de dispositifs d'accès à l'information géographique numérique sur le Web (IDG-plateforme de données ouvertes). Ce protocole a été testé sur un corpus de 43 dispositifs institutionnels français, avec un focus sur ceux concernant la thématique du suivi et de la gestion du trait de côte en France. Il convient de rappeler que les méthodes numériques employées constituent une première phase d'analyse souvent exploratoire permettant de dégager des tendances globales sur la thématique étudiée et non de faire émerger des modèles généraux (Plantin et Monnoyer-Smith, 2013 ; Ash et al., 2016). Elles doivent être accompagnées de méthodes d'analyse complémentaires comme l'analyse de contenu et l'analyse thématique des pages renvoyées par certains dispositifs numériques.

D'un point de vue des limites de notre approche méthodologique, nous avons été confrontés à une dynamique de la donnée (en volume, mise à jour, et supports) et à des logiciels évolutifs (langages, librairies, systèmes d'exploitation), qui demandent de plus en plus de ressources logistiques et mettent en difficulté l'emploi d'outils traditionnels tels que les suites logicielles traditionnelles. Le volume de données hypertextuelles traitées a souvent entraîné des difficultés pour stabiliser la base de données (MongoDB). En effet, certaines infrastructures comme la plateforme Data.gouv ou l'IDG Vigicrues contiennent près d'un million de pages à une profondeur de 2 clicks, ce qui peut affecter les capacités de stockage et de traitement. D'autres dispositifs au contenu dynamique de type « Content Management Systems », à l'instar de Dynaliti n'ont pas pu être moissonnés par le logiciel. Pour autant, si l'utilisation du caractère exploratoire d'un logiciel de « crawl » tel que Hyphe doit s'inscrire dans une problématique de recherche initiale, celui-ci peut aussi contribuer à la faire évoluer comme cela a été le cas dans la sélection de notre corpus qui a été redessiné au cours du projet de recherche.

#### 5. Conclusion

Le recensement des liens hypertextes, pages Web et sites Web renvoyés par chaque dispositif étudié révèle à ce jour une structure très hétérogène. En effet, un premier groupe de dispositifs renvoie à un faible nombre de liens, de pages et d'entités Web parmi lesquels figurent 5 dispositifs en lien avec l'information géographique numérique sur le trait de côte ( $n = 5$  dispositifs sur les 15 dispositifs en lien avec l'information géographique numérique sur le trait de côte présentent des valeurs inférieures à Q1). Ce sont le plus souvent des dispositifs spécialisés dans une forme de consultation unique de la donnée géographique (visualisation, génération de graphiques, catalogage de données). Ces infrastructures en ligne apparaissent dans les années 2005-2010 au moment de l'émergence des premiers dispositifs de cartographie grand public sur le Web (Mericksay, 2011 ; Noucher, 2017). D'autres au contraire sont très récents comme les applications de suivi du trait de côte par smartphone (information géographique volontaire). Il s'agit de dispositifs peu éditorialisés (faible quantité de liens et de pages Web au niveau hypertextuel) et peu

visibles dans le réseau des IDG institutionnelles (faible score d'autorité et peu interconnecté), au sein desquels le cheminement pour l'utilisateur est restreint à la fonction initiale du dispositif. Au contraire, certaines IDG nationales thématiques en lien avec les politiques publiques environnementales (biodiversité, eau, risque inondation) ainsi que la plateforme de données ouvertes nationale data.gouv sont plus visibles et offrent à l'utilisateur la possibilité d'élargir sa consultation à un grand nombre de ressources.

Enfin la composante réseau de ces dispositifs, deuxième indicateur structurel souvent mis en avant pour caractériser les IDG et autres plateformes de diffusion de données géographiques (Noucher et al., 2017), fait apparaître une connexité encore perfectible pour faciliter l'interopérabilité et l'accès aux différentes ressources offertes par l'ensemble des dispositifs. Ainsi, au sein des dispositifs identifiés par les pouvoirs publics dans la stratégie nationale de gestion intégrée du trait de côte, seules les IDG nationales Geolittoral et l'ONML se positionnent dans les scores les plus élevés en termes de centralités de réseaux (« Autorités », « Hubs », « PageRank »), tout en jouant des rôles différents dans les deux niveaux. Geolittoral, dispositif mis en avant dans la stratégie nationale de gestion intégrée du trait de côte, présente un profil de « hub » dans le niveau hypertextuel et d'autorité au niveau des services de données. A contrario, l'ONML dédié à une information plus globale (synthèses, rapports) et à un public plus large présente un profil d'autorité au niveau hypertextuel et de « hub » au niveau des services. Ces infrastructures exhibent donc une structure sur le Web que l'on peut qualifier de niveau unique : « hub » et autorité uniquement à un niveau de réseau (hypertexte ou services web) et services de consultation de la donnée d'un seul type. L'autre constat qui émerge de cette recherche d'un point de vue structurel est la présence quasi systématique de dispositifs généralistes, thématiques nationaux et territoriaux aux indices de structure hypertextuelle et aux niveaux d'interopérabilité les plus élevés, mais qui sont pourtant encore peu pointés dans la stratégie nationale de gestion intégrée du trait de côte française.

Ces résultats, qui concernent les données hypertextuelles et les services de consultation des dispositifs sur le Web, seront prochainement complétés par une analyse des données géographiques cataloguées par ces dispositifs en ligne et par des entretiens menés avec des promoteurs et des utilisateurs afin de mieux comprendre l'aspect structurel et la disponibilité de l'information géographique sur ces dispositifs<sup>16</sup>.

---

<sup>16</sup> Ce travail s'inscrit dans le cadre de la thèse AInfoGéo (<http://www-iuem.univ-brest.fr/pops/projects/ainfogeio>) dont l'objectif est de proposer un cadre méthodologique pour étudier la contribution de l'information géographique numérique disponible sur des dispositifs en ligne comme les IDG et plateformes de données ouvertes dans les politiques publiques en lien avec le suivi et la gestion du trait de côte.

## Bibliographie

- Ash J., Kitchin R. & Leszczynski A. (2016). Digital turn, digital geographies? *Progress in Human Geography*. vol. 42, p. 25-43, <https://doi.org/10.1177/0309132516664800>.
- Beaude B. (2012). *Internet, changer l'espace changer la société les logiques contemporaines de synchronization*, éditions Fyp, collection société de la connaissance.
- Beauguitte L. (2010). Graphes, réseaux, réseaux sociaux : vocabulaire et notation : Version 1 - Septembre 2010. Groupe f.m.r. (flux, matrices, réseaux) animé par César Ducruet et Laurent Beauguitte (CNRS, UMR , 2010 <halshs-00541898>)
- Beauguitte L. (2013). *D comme densité*, <https://groupefinr.hypotheses.org/2653>
- Bersani C., M.-L. Simoni, Y.-M. Allain, G. Ribière, J. Denègre, P. Planques, Y. Tugayé (2006). *Rapport relatif au schéma d'organisation des dispositifs de recueil de données et d'observation sur le littoral*, Ministère de l'intérieur et de l'aménagement du territoire, Ministère des transports, de l'équipement, du tourisme et de la mer, 119 p, [http://www.shom.fr/fileadmin/data/DOPS/MIP/PEP/BATHY/Litto3D/Besoins/Rapport-Bersani\\_recueil-donnees-littoral.pdf](http://www.shom.fr/fileadmin/data/DOPS/MIP/PEP/BATHY/Litto3D/Besoins/Rapport-Bersani_recueil-donnees-littoral.pdf)
- Cardon D. (2011). L'ordre du Web. *Médium*, 2011/4 vol. n° 29, p. 191-202. DOI: 10.3917/mediu.029.0191. URL: <https://www.cairn.info/revue-mEDIUM-2011-4-page-191.htm>
- Chakrabarti S. (2002). *Mining the Web: Discovering knowledge from hypertext data*, Elsevier.
- Colin N. et Verdier H. (2015). *L'âge de la multitude-2e éd.*, *Entreprendre et gouverner après la révolution numérique*. Armand Colin.
- Desbois H. (2015). *Les mesures du territoire : aspects techniques, politiques et culturels des mutations de la carte topographique*, Presses de l'ENSSIB, Nancy.
- Fruchterman, T. M. J., Reingold, E. M. (1991), Graph Drawing by Force Directed Placement, *Software: Practice and Experience*, vol. 21, no 11, p. 1129-1164.
- Gautreau, P., Severo, M., Giraud, T., Noucher, M. (2013), Formes et fonctions de la « donnée » dans trois webs environnementaux sud-américains (Argentine, Bolivie, Brésil), *Networks and Communication Studies (NETCOM)*, vol. 27 (2013), n° 1-2 pp. 22-59.
- Gasser U. (2015). *Interoperability in the Digital Ecosystem*, Berkman Klein Center for Internet and Society Research Publication No. 2015-13. Available at SSRN: <https://ssrn.com/abstract=2639210> or <http://dx.doi.org/10.2139/ssrn.2639210> <https://dash.harvard.edu/handle/1/28552584>
- Georis-Creuseveau. J. (2014). *Les Infrastructures de Données Géographiques (IDG) : développement d'une méthodologie pour l'étude des usages. Le cas des acteurs côtiers et de la GIZC en France*. Thèse de Géographie, Université de Bretagne Occidentale.
- Georis-Creuseveau J., Claramunt C., Gourmelon F. (2016). A modelling framework for the study of Spatial Data Infrastructures applied to coastal management and planning. *International Journal of Geographical Information Science*, vol. 31, no 1, p. 122-138. DOI: 10.1080/13658816.2016.1188929
- Girard P., Jacomy M., Baneyx A. (2012). The Hypertext Corpus Initiative: methods and tools for Social Sciences to build corpus from the Web, *Digital Humanities Luxembourg 2012*

- Websites as Sources, Luxembourg, [http://www.medialab.sciences-po.fr/publications/HypertextCorpusInitiative\\_webassources\\_abstract.pdf](http://www.medialab.sciences-po.fr/publications/HypertextCorpusInitiative_webassources_abstract.pdf).
- Graham M, and Zook M. (2013). Augmented Realities and Uneven Geographies: Exploring the Geo-Linguistic Contours of the Web. *Environment and Planning A* 45, n°1, p. 77–99.
- Graham M., & Dutton W. H. (Eds.) (2014). Introduction, *Society and the internet: How networks of information and communication are changing our lives*, OUP Oxford, p. 1-19.
- Jacomy M., Ghitalla F. (2007). *Méthodologie d'analyse de corpus en Sciences Humaines à l'aide du Navicrawler*. Rapport final, Fondation de la Maison des Sciences de l'Homme ProgrammeTIC-Migrations , <http://webatlas.fr/wp/share/navicrawler/Guide%20m%E9thodo%20NC%202007.pdf>.
- Kleinberg J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, vol. 46, no 5, p. 604-632.
- Mericksay, B. (2011). Les Sig et la cartographie à l'ère du géoweb Vers une nouvelle génération de Sig participatifs. *L'Espace géographique*, 2011/2 Tome 40, p. 142-153.
- Noucher M. (2013), Infrastructures de données géographiques et flux d'information environnementale, *Networks and Communication Studies (NETCOM)*, 27-1/2 | 2013, p. 120-147.
- Noucher M., Gourmelon F., Gautreau P., Georis-Creuseveau J., Maulpoix, A., Pierson J., Rouan M. (2017). Spatial Data Sharing: A Pilot Study of French SDI., *ISPRS International Journal of Geo-Information*, vol. 6, no 4, p. 99.
- Noucher M. (2017). *Les Petites Cartes du web. Approche critique des nouvelles fabriques cartographiques*, Editions Rue d'Ulm, collection Actes de la recherche à l'Ens n°18.
- Page, L., Brin, S., Motwani, R., Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the web*, Stanford InfoLab, <http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>.
- Pinède N. (2014). Le réseau, un « entre-deux » liens ? De quelques facettes et dynamiques du lien hypertexte. *Sciences de la société*, 91 | 2014, 94-105.
- Plantin J-C, Monnoyer-Smith L. (2014). Ouvrir la boîte à outils de la recherche numérique. *tic&société* [En ligne], Vol. 7, N° 2 | 2ème semestre 2013, mis en ligne le 01 juin 2014, Consulté le 02 juin 2014. URL : <https://journals.openedition.org/ticetsociete/1527> ; DOI : 10.4000/ticetsociete.1527.
- Pornon H. (2015). *SIG : La dimension géographique du système d'information (2ème édition)*, Management des systèmes d'information, Paris, Dunod.
- Rajabifard, A., Feeney, M.-E., Williamson, I.P. (2002). Future directions for SDI development. *International Journal of Applied Earth Observation and Geoinformation*, vol. 4, n° 1, p. 11-22.
- Rogers R. (2009). Situating Digital Methods in Internet Research, *The End of the Virtual: Digital Methods*, Amsterdam University Press.
- Rogers, R. (2015). Digital Methods for Web Research, *Emerging Trends in the Social and Behavioral Sciences* (eds R. A. Scott and S. M. Kosslyn). doi:10.1002/9781118900772.etrds0076.

- Severo M., Venturini T. (2016). Enjeux topologiques et topographiques de la cartographie du Web. Le cas du patrimoine culturel immatériel français. *Réseaux*, 1/2016 n° 195, p. 85-105.
- Suarez S., Garcin M., Bulteau T., Rouan M., Lagadec L. & David L. (2012). Les observatoires du trait de côte en France métropolitaine et dans les DOM. *EchoGéo* [En ligne], 19 | 2012, mis en ligne le 10 février 2012, consulté le 12 décembre 2016. URL : <http://journals.openedition.org/echogeo/12942> ; DOI : 10.4000/echogeo.12942.

---

## Vers un observatoire des données géographiques du Web.

### Expérimentation d'une analyse diachronique des infrastructures de données géographiques

**Matthieu Noucher<sup>1</sup>, Françoise Gourmelon<sup>2</sup>, Grégoire Le Champion<sup>1</sup>, Adeline Maulpoix<sup>2</sup>, Julie Pierson<sup>1</sup>, Olivier Pissot<sup>1</sup>, Mathias Rouan<sup>2</sup>**

1. CNRS - Laboratoire PASSAGES (UMR 5319) - Bordeaux  
Maison des Suds, 12 esplanade des Antilles, 33607 Pessac, France

2. CNRS - Laboratoire LETG (UMR 6554) - Brest  
IUEM, rue Dumont d'Urville, 29280 Plouzané, France

Auteur correspondant : [matthieu.noucher@cnrs.fr](mailto:matthieu.noucher@cnrs.fr)

---

*RESUME.* L'extension des usages de l'information géographique génère une profusion de données qui circulent sur de multiples plateformes Web. Accompagner le suivi du contenu de ces plateformes relève aujourd'hui de la gageure tant elles s'apparentent à de véritables boîtes noires algorithmiques et que leurs dynamiques sont élevées. Fort de ce constat, il apparaît que le concept d'observatoire de l'information géographique est intéressant pour suivre, analyser et comprendre l'évolution du volume et des flux de données géographiques disponibles sur une portion de territoire. Néanmoins sa mise en œuvre soulève plusieurs enjeux méthodologiques que nous discutons à partir de l'expérimentation d'une observation diachronique du patrimoine de données issues des Infrastructures de Données Géographiques (IDG) françaises.

*ABSTRACT.* Extending uses of geographic information generates a data deluge on Web platforms. Monitoring the content of these platforms is a challenge because they are similar to algorithmic black boxes and their dynamics are very high. Based on this observation, the concept of "geographical information observatory" appears interesting to monitor, analyze and understand the evolution of the volume and the flow of spatial data available on a specific place. Nevertheless, its implementation raises several methodological issues that we discuss starting from the experimentation of a diachronic observation of data from French Spatial Data Infrastructures (SDI).

*MOTS-CLES :* Infrastructure de Données Géographiques (IDG) ; Observatoire de l'Information Géographique ; Analyse réseau ; Analyse spatiale ; Analyse diachronique.

*KEYWORDS:* Spatial Data infrastructure (SDI) ; Geographic Information Observatory ; Network analysis ; Spatial analysis ; Diachronic analysis.

---

## 1. Introduction

Au-delà des raccourcis sur le « data deluge » ou la « révolution géospatiale »<sup>1</sup> l'un des enjeux de recherche actuel en géomatique est d'être en capacité de dresser un panorama des données géographiques pour mieux comprendre à la fois les opportunités offertes par des sources devenues multiples mais aussi les lacunes informationnelles sur certains territoires ou certaines thématiques.

Fort de ce constat, deux workshops ont été organisés lors de la 8<sup>e</sup> conférence GIScience à Vienne (Autriche) en 2014 et de la conférence COSIT à Santa Fe (USA) en 2015, autour du concept des *Geographic Information Observatories*. La nécessité d'explorer *l'univers de l'information géographique* en déployant des protocoles d'observation qui permettent un suivi dans le temps des données en ligne, en termes de contenus et d'usages a été soulignée par Ballatore (2014). Alors que l'information géographique est régulièrement mobilisée dans les phases de collecte, de traitement et de restitution des indicateurs issus des observatoires territoriaux (de Sède-Marceau et al., 2018), le développement du géoweb permet aujourd'hui d'envisager la mise en place d'observatoires de l'information géographique pour évaluer sur un territoire et/ou une thématique spécifique l'état des lieux des données géographiques disponibles. À l'instar des observatoires scientifiques qui ont pour objectif « d'observer (suivre, analyser, comprendre), sur une portion d'espace représentative d'une entité territoriale, les dynamiques socio-environnementales résultant d'interactions entre systèmes socio-économiques et biophysiques » (Libourel et al., 2009), les observatoires de l'information géographique sur le Web peuvent se définir comme des dispositifs sociotechniques en réseau mis en œuvre pour suivre, analyser et comprendre l'évolution spatio-temporelle des données géographiques et de leurs usages sur un territoire pour en révéler les représentations spatiales associées à ses dynamiques socio-environnementales. Lieu de production, d'échange et de partage d'informations et de connaissances qui s'inscrit dans la durée (Janowicz et al., 2014), ces observatoires se concentrent sur un contenu numérique et auraient pour principales fonctions de rendre visible et de mémoriser un patrimoine immatériel difficile à saisir et parfois souvent évanescent car malléable.

De notre point de vue, la mise en œuvre d'observatoires de l'information géographique soulève au moins trois questions méthodologiques : 1) comment, face à des contenus qui circulent sur différentes plateformes Web être en mesure d'identifier les données géographiques accessibles ? 2) une fois identifiées, comment les qualifier ? 3) enfin, comment en assurer le suivi dans le temps pour les voir évoluer. Ce dernier point nous semble particulièrement important puisque l'ambition de déploiement d'un observatoire de l'information géographique ne peut se concevoir comme une simple photographie à un instant « t » mais bien comme un

---

<sup>1</sup> Nous reprenons ici l'intitulé d'un web-documentaire de l'Université Penn State (USA) dont le titre et le contenu illustrent bien le discours (caricatural) de rupture qui accompagne la promotion de ces développements technologiques : <http://geospatialrevolution.psu.edu/>



dispositif d'accompagnement – c'est-à-dire en lien direct avec les observatoires et plateformes, notamment institutionnelles, existantes – qui s'inscrit dans la durée.

Dans cette contribution, nous proposons d'expérimenter un protocole d'identification, de qualification et de suivi de l'évolution des données issues des Infrastructures de Données Géographiques (IDG) françaises. Les IDG sont, en effet, devenues des plateformes web essentielles pour la diffusion des données institutionnelles (Rajabifard et al., 2003). Elles visent à faciliter, à tous les échelons territoriaux, la découverte, l'accès, l'échange et le partage d'information et de services géographiques (Masser, 2010). Les IDG contribuent à la complexité de *l'univers de l'information géographique* (Ballatore, 2014 ; Janowicz et al., 2014) et sont considérées par certains auteurs (Grus et al., 2010) comme des systèmes adaptatifs qui ajustent leur contenu, leurs actions et objectifs en fonction de facteurs externes.

Notre corpus est constitué de 39 IDG nationales et régionales qui disposent d'un géocatalogue opérationnel. A travers leur exploration, nous produisons des indicateurs concernant leur couverture spatiale, thématique ou encore organisationnelle sur 3 ans (2016 à 2018). Ce faisant, cette expérimentation peut s'apparenter à une tentative de prototypage d'un observatoire de l'information géographique issue des IDG dont nous soulignons, en conclusion, les enjeux et limites.

## 2. Méthodologie

### 2.1. Extraction des données

Pour extraire les informations contenues dans les métadonnées des 39 IDG étudiées<sup>2</sup>, un script Python<sup>3</sup> 2.7 utilisant la bibliothèque OWSLib<sup>4</sup> a été développé (Rouan et Pierson, 2016). Cette bibliothèque offre la possibilité d'interroger des catalogues en utilisant les différents standards de l'Open Geospatial Consortium (OGC), dont le service web de catalogage (CSW) dans sa version 2.0.2. Pour cette phase d'extraction, plusieurs problèmes d'ordre technique ont été rencontrés. Selon la plate-forme utilisée par l'IDG (GeoNetwork, Isogeo, Prodige, Amigo...), l'implémentation du CSW peut varier ; une IDG pouvant ne plus être interrogeable (ré-indexation, problème réseau) pendant l'exécution du script. Il faut donc pouvoir relancer régulièrement la procédure. Les CSW de notre corpus ont donc été testés sur une période de plusieurs semaines à chaque fois : en juillet pour l'extraction de

<sup>2</sup> Echantillon établi à partir d'une sélection des IDG régionales et nationales, inventoriées dans l'annuaire de l'AFIGEO et disposant d'un géocatalogue opérationnel.

<sup>3</sup> <https://github.com/LETG/csw-harvester>

<sup>4</sup> <https://github.com/geopython/OWSLib>

2016 et en mai celle de 2017<sup>5</sup>. D'autres problèmes inhérents aux contenus des fiches de métadonnées ont été identifiés. Par exemple toutes les rubriques, même celles rendues obligatoires par la norme ISO 19115, ne sont pas toujours remplies ou peuvent être mal renseignées. C'est le cas des dates (création, publication, révision) ou encore de la généalogie dont le champ de saisie libre s'est avéré particulièrement mal renseigné pour permettre une analyse systématique.

La librairie OWSLib a permis de récupérer les fichiers XML des métadonnées et de réaliser leur analyse avec des requêtes XQuery. Cependant le souhait de rendre ces données plus accessibles notamment en proposant des analyses et des synthèses dynamiques à travers le Web nous a conduit à les structurer dans une base de données relationnelle. Cette phase d'archivage est réalisée au moyen de la bibliothèque Psycopg<sup>6</sup>, adaptateur PostgreSQL le plus utilisé en Python.

## 2.2. Archivage et gestion temporelle des données

Concernant l'archivage des données, l'objectif n'est pas d'explorer l'ensemble des métadonnées mais de les analyser au regard de quelques questions de recherche (Noucher et al., 2017). Il n'était donc pas question de proposer un schéma de base de données reprenant l'ensemble de la norme. Dans cette perspective, une sélection d'éléments pertinents a été réalisée et un modèle entité-association élaboré. Celui-ci intègre une dimension temporelle en relevant pour chaque opération, une date d'extraction (figure 1).

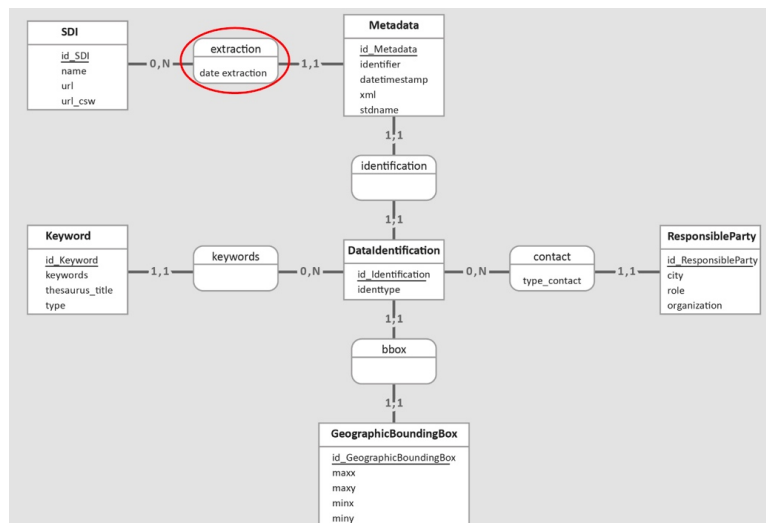


Figure 1. Modèle conceptuel de données (MCD).

<sup>5</sup> Une autre extraction est prévue avant l'été 2018. Les résultats présentés lors de la conférence SAGEO seront enrichis de cette nouvelle analyse.

<sup>6</sup> <http://initd.org/psycopg/>

Afin de rendre la base de données évolutive, nous avons fait le choix de garder la structuration et le nom des tables/champs de la norme. Ainsi, le MCD contient 3 entités (« Metadata », « DataIdentification » et « GeographicBoundingBox ») qui sont associées par des relations de cardinalité 1,1 ce qui occasionne une multiplication injustifiée du nombre de tables et donc une complexification des requêtes par l'ajout de jointure. Cependant cette structuration a l'avantage d'être évolutive en permettant la création de nouveaux attributs ou entités/rerelations tout en conservant l'intégrité de la base sans modifier les requêtes existantes. Le choix a également été de conserver l'ensemble des métadonnées en XML dans un champ de la table « Metadata » permettant ainsi de ne perdre aucune information qui pourrait être utile pour re-peupler la base en cas d'évolution. La base de données a été implémentée sous PostgreSQL 9.3.

### **2.3. Choix des indicateurs de suivi**

Le prototype d'un observatoire des géocatalogues des IDG françaises que nous proposons repose sur des indicateurs qui visent à interroger quatre « apports » des IDG en termes de mise en réseau (des données, des outils, des acteurs, des territoires) souvent mis en avant par leurs promoteurs. Plusieurs méthodes d'analyse (spatiale, statistique, réseau) ont été mobilisées pour produire ces indicateurs. Elles seront présentées ultérieurement en association avec chaque type de résultat.

A titre illustratif nous en proposons quatre (tableau 1) qui ont fait l'objet d'une restitution sous la forme de (géo)visualisations (cartes, graphes, graphiques) pour l'année 2016 (Noucher et al., 2017). Cependant, comme évoqué en introduction, l'un des principaux enjeux des observatoires de l'information géographique est de dépasser la situation à un instant « t » du patrimoine de données pour le suivre dans le temps et évaluer son évolution.

Tableau 1. Exemples de fiches méthodologiques pour 4 indicateurs.

	Mise en réseau des DONNEES	Mise en réseau des OUTILS	Mise en réseau des ACTEURS	Mise en réseau des TERRITOIRES
<b>Présentation générale</b>				
Titre indicateur	Accessibilité des données	Interopérabilité des catalogues	Couverture organisat <sup>o</sup> des données	Couverture spatiale des données
Définition	Part des données téléchargeables et réutilisables sans condition.	Interconnexion des géocatalogues des IDG et des autres plateformes Web.	Part des différents types de contributeurs des géocatalogues	Distribution géographique de l'emprise des données diffusées
Objectifs	<i>Visualiser et suivre l'évolution de la proportion de données effectivement utilisables pour dépasser le comptage des métadonnées.</i>	<i>Visualiser et suivre l'évolution du réseau de moissonnage pour comprendre comment circule la donnée.</i>	<i>Visualiser et suivre l'évolution des types d'organisation (autorités publiques, sociétés privées...) qui alimentent les géocatalogues.</i>	<i>Visualiser et suivre l'évolution de la densité de données disponibles sur les territoires (hots spots informat<sup>o</sup>, zone blanches...).</i>
Données en entrée	Métadonnées des géocatalogues.	Liens moissonneurs / moissonnés	Métadonnées des géocatalogues.	Métadonnées des géocatalogues.
Sources	Service web CSW	Enquête en ligne	Services web CSW	Services web CSW
<b>Modalité de calcul</b>				
Traitement statistiques	Comptage et pourcentage	Analyse réseaux	Comptage et pourcentage	/
Traitement géométriques	/	/	/	Analyse spatiale
Traitement sémantiques	Filtre sur tags <i>opendata / donnée ouverte</i>	/	Catégorisation dans une typologie <i>ad hoc</i>	/
<b>Processus d'actualisation</b>				
Fréquence	Annuelle	Bi-annuelle	Annuelle	Annuelle
Méthode	Relance du script	Enquête en ligne	Relance du script	Relance du script
Source	Service web CSW	Promoteur IDG	Service web CSW	Service web CSW
<b>Interprétation des résultats</b>				
Comparaison des résultats	Analyse diachronique Comparaison volume de métadonnées VS volume de données dites ouvertes	Analyse diachronique	Analyse diachronique	Analyse diachronique
Echelle d'analyse	Analyse globale, par échelon (régional, national), par IDG	Analyse globale sur l'ensemble du réseau des IDG	Analyse globale, par échelon (régional, national), par IDG	Analyse globale, par échelon (régional, national), par IDG
Pondération des résultats	A pondérer avec la part des URL opérationnels	/	/	/
Combinaison possible	Avec l'ensemble des autres indicateurs	/	Avec l'ensemble des autres indicateurs	Avec l'ensemble des autres indicateurs
<b>Représentation</b>				
Tableau	Comptage Pourcentage	/	Comptage Pourcentage	/
Graphiques	Courbe d'évolution	/	Diagramme circulaire	/
Cartes	Distribution de la densité des données	/	/	Cartes des hots spots informationnels
Autres	/	Graphe	/	/

### 3. Résultats

Naviguer sur les sites web des Infrastructures de Données Géographiques ne permet pas d'identifier de manière exhaustive le patrimoine de données disponible en ligne. Les interfaces de navigation des géoportails n'offrent souvent que des possibilités fonctionnelles limitées pour visualiser un nombre de couches restreint. Par ailleurs, l'entrée sur les géocatalogues par des moteurs de recherche ne donne accès qu'à un ensemble filtré (en fonction des paramètres des requêtes) de métadonnées. Présentés sous la forme de liste, dont l'ordonnancement est d'ailleurs rarement explicité, les résultats obtenus *via* ces moteurs de recherche n'offrent aucune possibilité de post-traitement pour explorer ces corpus. Quelques tableaux de bord (*dashboard*) existent depuis peu. Le BRGM et Etalab en proposent depuis 2017. Mais ces derniers ne semblent s'inscrire que dans une logique d'évaluation. Ils sont ainsi centrés sur des cibles restreintes : uniquement les données géographiques qui sont dans le périmètre de la Directive INSPIRE pour le dashboard du BRGM<sup>7</sup>, uniquement les données en opendata pour celui d'Etalab<sup>8</sup>. Par ailleurs, en fournissant un panorama ponctuel des métadonnées, ces tableaux de bord n'assurent pas l'archivage et l'analyse des indicateurs produits. Si certains indicateurs sont répétés avec des fréquences régulières (tous les 3 ans dans le cadre du rapportage de la Directive INSPIRE, par exemple), les méthodes semblent loin d'être stabilisées :

*« Cette évolution peut être relativisée par une modification de la méthodologie appliquée cette année dans le calcul de cet indicateur. Jusqu'à 2015, seul le premier thème renseigné dans les métadonnées était pris en compte pour calculer le nombre de métadonnées de données. En 2016, les séries de données sont comptées autant de fois qu'elles référencent dans leurs métadonnées de thèmes INSPIRE, augmentant ainsi artificiellement le nombre de métadonnées de données. A titre de comparaison, le nombre de métadonnées de données passe de 25501 (selon l'ancienne méthode) à 30452 (selon la nouvelle méthode). »*

A propos de l'indicateur MDi1, p.4-5 du « Rapport de la France sur la mise en œuvre de la directive INSPIRE pour la période 2013-2015 (réf. 2013-2015) »

Pour dépasser ces limites, la répétition selon les mêmes protocoles des opérations d'extraction et d'archivage (2.1 et 2.2) nous a permis d'appliquer les mêmes méthodes de qualification (2.3) des métadonnées, permettant *de facto* une analyse diachronique de nos quatre indicateurs de mise en réseau.

<sup>7</sup> <http://dashboard.geocatalogue.fr/daobs/dashboard/>

<sup>8</sup> <https://www.data.gouv.fr/fr/dashboard/>

### 3.1. Mise en réseau des données : évolution de l'accessibilité des contenus

L'analyse des données accessibles (téléchargeables et réutilisables sans contrainte) au sein des IDG permet de nuancer le discours sur le « déluge de données ». Si l'indexation de métadonnées connaît une croissance exponentielle, les données réellement accessibles restent minoritaires (tableau 2).

Tableau 2. Evolution des volumes de métadonnées et de la part des données accessibles dans les IDG entre 2016 et 2017.

	Juin 2016	Mai 2017	Différentiel
Volume total de métadonnées	160603	188602	+17 %
Moyenne de métadonnées par IDG	4340	5097	+17 %
Moyenne des métadonnées dans les IDG nationales	7159	9303	+30 %
Moyenne des métadonnées dans les IDG régionales	2987	3317	+11%
Part des « données ouvertes » / « open data »	15,7 %	19,1%	+3,4 pts
Part moyenne des données ouvertes dans les IDG nationales	8,5%	22,7%	+13,2 pts
Part moyenne des données ouvertes dans les IDG régionales	19,1%	10,5%	- 8,6 pts

L'analyse de l'évolution 2016-17 prend ici tout son sens puisqu'elle met en évidence des soldes négatifs qui s'expliquent par la fusion des régions<sup>9</sup> et qui a conduit à certaines reconfigurations ou au report de données sur des portails *open data*.

### 3.2. Mise en réseau des outils : évolution de l'interopérabilité des géocatalogues

Pour évaluer l'interconnexion des IDG françaises, nous réalisons une analyse réseau du moissonnage de leurs géocatalogues. Les graphes qui en résultent pour 2015 et 2017 (figure 2) permettent d'identifier les nœuds centraux (ou isolés) et d'éventuels *clusters*. Des composantes géographiques similaires aux deux années sont mises en évidence à travers la relation des IDG régionales aux plateformes infrarégionales. En l'occurrence, les jeux d'échelle sont très marqués à travers la structure centre-périphérie du graphe. La plupart des IDG nationales occupe une place centrale. Une première couronne périphérique donne un panorama des IDG régionales qui renvoient elles-mêmes vers une deuxième couronne périphérique composée majoritairement d'IDG initiées par des collectivités territoriales, des associations ou des syndicats départementaux ou intercommunaux. De même, des composantes thématiques du réseau apparaissent à la fois en 2015 et en 2017 à travers les mises en relation d'IDG travaillant dans le domaine de la mer ou du côtier et sur des territoires ultramarins. Enfin, certaines infrastructures utilisant peu le moissonnage (nœuds de petite taille) occupent pourtant, depuis 2015, une place centrale car elles sont très moissonnées. C'est le cas d'EauFrance et de Géo-IDE.

<sup>9</sup> Loi NOTRe du 7 août 2015 portant nouvelle organisation territoriale de la République

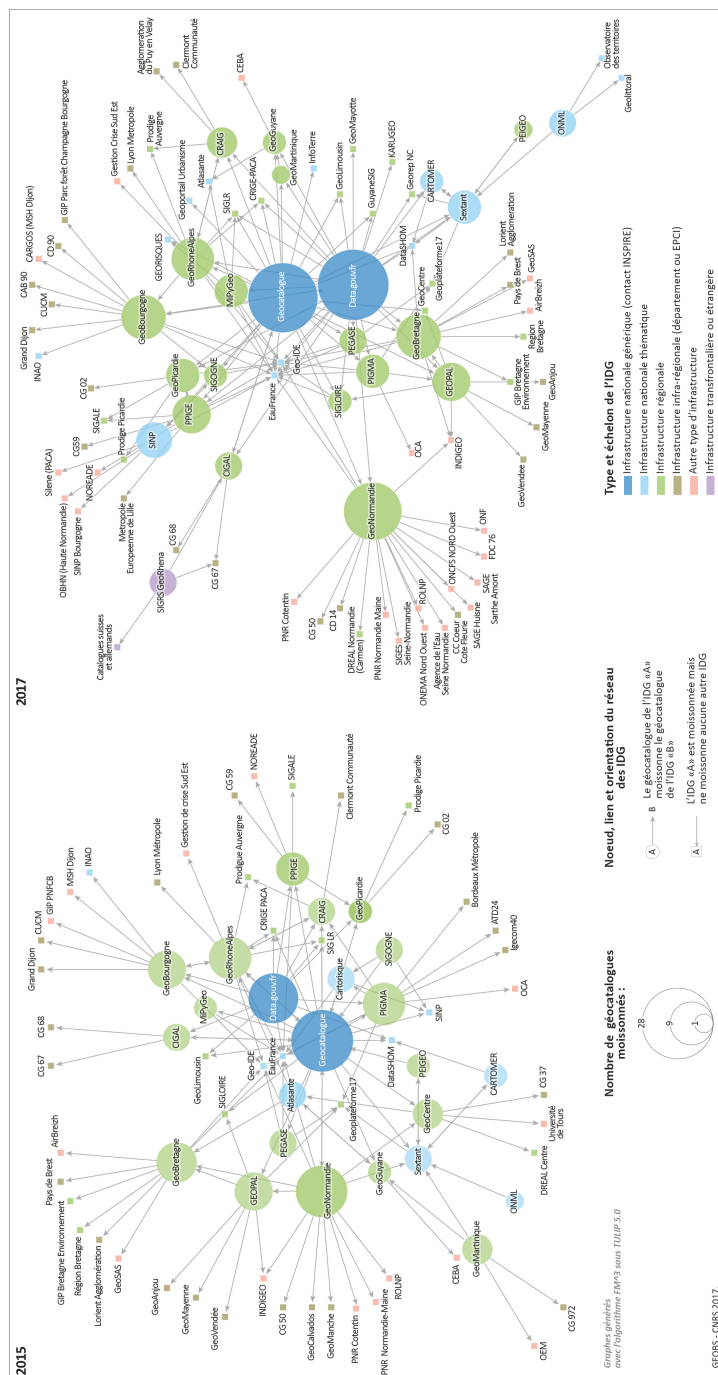


Figure 2. Evolution du réseau de moissonnage entre 2015 et 2017

Si les réseaux de moissonnage de 2015 et 2017 partagent des caractéristiques communes<sup>10</sup>, l'analyse diachronique permet de pointer quelques évolutions. On peut ainsi noter en 2017 l'apparition dans le réseau d'IDG étrangères par l'intermédiaire de géocatalogues transfrontaliers ainsi que l'extension du moissonnage vers des systèmes infra-départementaux (intercommunalités essentiellement). La distribution de l'influence dans le réseau a évolué. L'indicateur moyen de *centralité d'intermédierité*<sup>11</sup> a décliné entre 2015 et 2017 en raison d'une forte baisse de cet indicateur pour le Géocatalogue national. Si celui-ci apparaissait comme le principal *hub* (nœud ayant le plus fort degré entrant) en 2015 (21 moissons), il est dépassé en 2017 par le portail data.gouv.fr (28 moissons). Cette évolution traduit bien les reconfigurations en cours entre IDG et portails *open data*.

### 3.3. Mise en réseau des acteurs : évolution de la couverture organisationnelle

L'analyse des métadonnées révèle également la très faible diversité des contributeurs des IDG (figure 3). Sur les plus de 230000 contacts recensés dans les métadonnées de 2016, 70% proviennent des autorités publiques et seulement 3% du secteur privé. Cette proportion a faiblement évolué en 2017 malgré les discours volontaristes sur l'ouverture de ces infrastructures professionnelles vers l'information géographique volontaire (VGI). En revanche, la part des données issues de l'enseignement et de la recherche a fortement augmenté (de 11,7 à 15,9%).

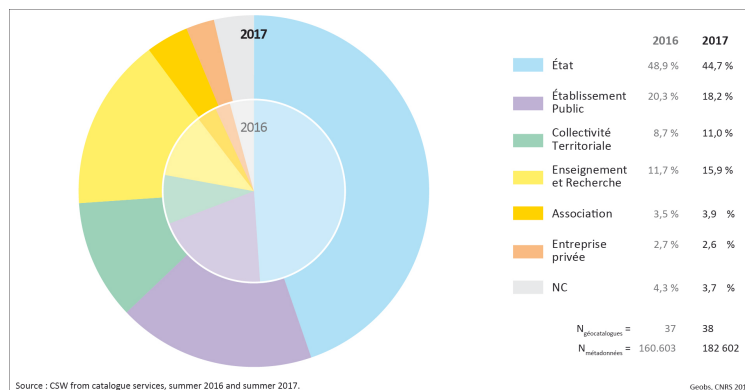


Figure 3. Evolution de la couverture organisationnelle des géocatalogues.

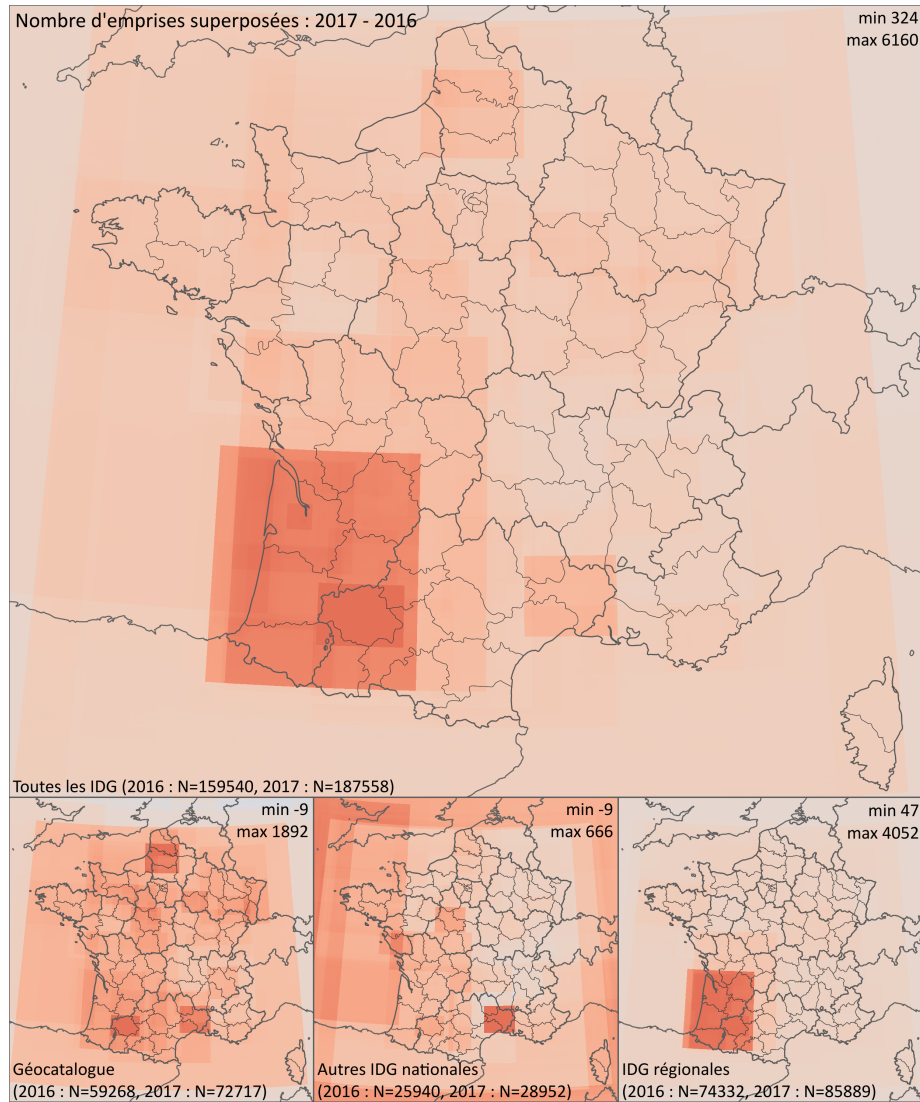
### 3.4. Mise en réseau des territoires : évolution de la couverture spatiale

La cartographie de l'emprise spatiale des données interroge l'égalité informationnelle des territoires (Gautreau et Noucher, 2016).

<sup>10</sup> Par exemple, la *densité*  $d$  qui mesure la complétude d'un graphe ( $d=1$  si tous les liens possibles existent) est quasi similaire :  $d_{(2016)}=0,022$  et  $d_{(2017)}=0,02$ .

<sup>11</sup> La centralité d'intermédierité (*betweenness centrality*)  $C_B$  mesure la fréquence d'apparition d'un nœud sur les plus courts chemins entre les différents nœuds du réseau.  $C_{B(2016)}=23,99$   $C_{B(2017)}=6,33$





#### A propos des données

Les emprises de toutes les données ont été extraites des IDG en utilisant les CSW, en 2016 et 2017. Le nombre de superpositions de ces emprises est ensuite calculé pour chaque année. Le résultat obtenu pour 2016 est ensuite soustrait de celui obtenu pour 2017.

#### Discretisation : intervalles égaux

Sources : CSW-2 des services de catalogage, 2016 et 2017 - Natural Earth - GEOFLA®IGN

#### Nb superpositions

min à -n  
-n à 0  
0 à n  
n à 2n  
2n à 3n  
3n à 4n  
4n à 5n  
5n à max

Figure 4. Evolution de la couverture spatiale des métadonnées entre 2016 et 2017.

Au-delà de l'identification des *hot-spots* informationnels ou au contraire des *zones blanches* en données, il est aussi intéressant de suivre l'évolution des dynamiques de diffusion des données. La figure 4 présente ainsi le différentiel entre les couvertures spatiales des 160603 métadonnées recensées en 2016 et celles des 188602 métadonnées indexées<sup>12</sup> dans les mêmes géocatalogues en 2017. Si ce différentiel est quasi exclusivement positif, la discrétisation par quantile laisse entrevoir une certaine hétérogénéité dans l'augmentation du nombre de métadonnées. Les cartes présentées ici laissent ainsi apparaître l'importance de certaines initiatives départementales (Gers, Gard, Somme...) ou métropolitaines (Bordeaux) et la persistance des anciennes régions (l'Aquitaine plutôt que la Nouvelle Aquitaine) dans les dynamiques d'alimentation récente (sur 1 an) des infrastructures françaises de données géographiques.

#### 4. Conclusion et perspectives

Le Ministère de la Transition écologique et solidaire utilise le Géocatalogue national pour réaliser le rapportage<sup>13</sup> exigé par la Commission européenne dans le cadre de la Directive INSPIRE (décision n°2009/442/CE). Il s'appuie ainsi sur le BRGM qui produit un suivi annuel et un rapport triennal. Le dernier rapport triennal, validé par le CNIG, diffusé en juillet 2016, dresse le bilan de la période 2013-2015 et a été établi à partir d'une enquête, de retours d'utilisateurs et d'une série d'indicateurs. Ces derniers sont produits *via* l'analyse des séries de données et de services géographiques qui s'inscrivent dans le périmètre de la Directive et étaient disponibles au 31 décembre 2015 dans le géocatalogue national. Ils concernent essentiellement les suivis des métadonnées, des données, des services en réseaux et des actions de coordination. L'analyse de l'évolution des volumes (données, métadonnées, services) et de leur niveau de conformité permet de comparer entre elles les plateformes nationales et d'opérer leur suivi dans le temps. Le rapport de la France souligne ainsi le doublement annuel de métadonnées<sup>14</sup> et précise que près de 90% d'entre-elles sont conformes à la réglementation européenne. Mais ce rapport ne fournit aucune analyse précise du patrimoine de données disponible, des champs thématiques prioritaires, des territoires les plus / les moins couverts, des échelles des données diffusées, de la répartition par type d'organisation, des périodes de validité des données proposées. La décision de la Commission (2009/442/CE) du 5 juin 2009 inscrit le rapportage avant tout dans une logique de benchmarking<sup>15</sup> visant à créer un référentiel cognitif commun destiné à rendre les données comparables d'un

<sup>12</sup> Le nombre d'emprises traitées sur ces cartes est légèrement inférieur (159540 en 2016 et 187558 en 2017) au nombre de métadonnées indexées bien que plusieurs emprises puissent être associées à une métadonnée. Cette différence est liée au choix que nous avons fait, r pour faciliter la diffusion des résultats, de ne centrer ces cartes que sur l'hexagone.

<sup>13</sup> Rapport d'évaluation normé selon un modèle proposé par la Commission.

<sup>14</sup> De 4026 métadonnées en 2013 à 34473 en 2016 soit une augmentation de 8,6 fois.

<sup>15</sup> Le *benchmarking* est une technique issue du marketing pour comparer les performances des entreprises. Un *benchmark* est un indicateur chiffré de performance dans un domaine donné tiré de l'observation des résultats de l'entreprise qui a réussi le mieux dans ce domaine. Cet indicateur peut servir à définir les objectifs de l'entreprise qui cherche à rivaliser avec elle.

pays à l'autre, sans porter de réflexion de fond sur les types de contenus à créer ni sur leur pertinence. En ce sens, l'évaluation des IDG relève aujourd'hui davantage d'une « logique de contrôle et de stimulation des Etats en situant leurs performances sur des échelles normatives » (Alphandéry et al., 2012) que d'une volonté de mieux comprendre les connaissances territoriales qu'elles sont susceptibles de diffuser ou des flux informationnels qu'elles sont potentiellement en train de ré-agencer (Noucher, 2013).

Par cette contribution nous souhaitons démontrer l'intérêt – au-delà de ces logiques d'évaluation – de la mise en œuvre d'observatoires de l'information géographique circulant sur le Web pour appréhender les dynamiques qui affectent le capital informationnel des territoires et les fluctuations de sa répartition (thématique, spatiale, organisationnelle). Notre expérimentation, menée à partir de l'observation diachronique du contenu de 39 IDG françaises, nous a permis de tester, sur des volumes de données importants, l'intérêt de ce type d'analyse pour compléter le rapportage tel qu'il est réalisé actuellement en France. La poursuite de ce travail pour engager des analyses diachroniques sur un pas de temps plus long est envisagée. Néanmoins il est important de souligner deux limites à ce travail. Premièrement le caractère cumulatif des corpus disponibles en ligne interroge la manière de gérer les données obsolètes, les données mortes ou encore les données publiées par mégarde qui ont tendance à s'accumuler dans ces infrastructures rarement mises à jour. À ce stade, les observations diachroniques les agrègent sans pouvoir les distinguer, introduisant un biais dans les résultats. Deuxièmement, la faisabilité et la robustesse des analyses proposées sont liées aux caractères institutionnel et ouvert des IDG étudiées. La mise en œuvre d'observatoires de l'information géographique circulant sur le Web au sens de Janowicz et al. (2014) nécessiterait d'analyser d'autres types de plateformes moins normées (de cartographie collaborative ou issues des réseaux sociaux) qui diffusent des données géographiques plus difficilement accessibles. Leur prise en compte pour dresser un tableau de bord qui tend vers l'exhaustif nécessiterait de démultiplier les chaînes d'extraction et d'archivage pour s'adapter à leur spécificité ou détourner leur restriction d'accès.

L'approfondissement de l'analyse du patrimoine informationnel des IDG et la diversification des types d'infrastructures web à étudier représentent ainsi des défis de recherche pour opérationnaliser le concept d'observatoire de l'information géographique.

## Bibliographie

- Adams B., Gahegan M., Gupta P., Hosking R., (2014). Geographic Information Observatories for supporting Sciences. In Proceedings of Workshop on Geographic Information Observatories 2014, <http://ceur-ws.org/Vol-1273/>, 32-39
- Alphandéry P., Fortier A. et Sourdril A. (2012), Les données entre normalisation et territoire : la construction de la trame verte et bleue, *Développement durable et territoires*, 3/2.

- Ballatore A. (2014). Exploring the geographic information universe: the role of search technologies. *CEUR Workshop Proceedings 1273, Workshop on Geographic Information Observatories at GIScience*, Vienna, October 2014, pp. 23–28
- Gautreau P. et Noucher M. (2016). « Information géographique numérique et justice spatiale : les promesses du partage ». *Justice Spatiale / Spatial Justice*, n°10, <http://www.jssj.org>
- Grus L., Crompvoets J., Bregt A.K. (2010). Spatial data infrastructures as complex adaptive systems. *International Journal of Geographical Information Science*, 2010, 24, 439-463
- Janowicz K., Adams B., McKenzie G., Kauppinen T., (2014). “Towards geographic information observatories”. In *Proceedings of Workshop on Geographic Information Observatories 2014*, <http://ceur-ws.org/Vol-1273/>, 1-5
- Libourel T., Passouant M., Loireau M., (2009). Introduction du chapitre « Systèmes d’observation et observatoires ». *Dossier Agropolis International* 9, 27 p.
- Masser, I. (2010). *Building European Spatial Data Infrastructures*. ESRI Press: Redlands, CA, USA.
- Noucher M., Gourmelon F., Gautreau P., Georis-Creuseveau J., Maulpoix A., Pierson J., Pinède N., Pissot O. et Rouan M. (2017). Spatial Data Sharing: A Pilot Study of French SDIs. *ISPRS International Journal Geographic Information*. vol. 6, 99.
- Noucher M. (2013) « Infrastructures de données géographiques et flux d’information environnementale », *Netcom*, 27-1/2 | 2013, 120-147.
- De Sède-Marceau M.-H., François S., Pauc B. (2018). « OPTTEER, une expérience de construction collective d’un dispositif de connaissance et d’analyse territoriale par et pour les acteurs de la transition énergétique ». *Revue Internationale de Géomatique*, vol. 28/1.
- Rajabifard A., Feeney M.-E., Williamson I. et Masser, I. (2003). Chapter 6, National SDI Initiatives. In Williamson, I., Rajabifard, A., Feeney, M.-E.F., (Eds). *Development of Spatial Data Infrastructures: From Concept to Reality*. Taylor & Francis: London, UK, pp. 95–109
- Rouan M. et Pierson J. (2016). « CSW\_Harvester : une chaîne d'extraction et d'analyse d'informations à partir des services web de catalogage (CSW) des Infrastructures de Données Géographiques nationales et régionales en France ». *Atelier - EXtraction de Connaissances à partir de donnÉes Spatialisées (EXCES) – SAGEO'2016*. Nice, France.

---

## Référencement spatial indirect : modélisation à base de relations et d'objets spatiaux vagues

Mattia Bunel<sup>1</sup>, Ana-Maria Olteanu-Raimond<sup>1</sup>, Cécile Duchêne<sup>1</sup>

<sup>1</sup>Univ. Paris-Est, LASTIG COGIT, IGN, ENSG, F-94160 Saint-Mandé, France  
[prenom.nom@ign.fr](mailto:prenom.nom@ign.fr)

---

**RÉSUMÉ.** Cet article propose une méthodologie destinée à transformer une description de position exprimée dans un référentiel indirect en une position absolue dans un référentiel direct. Cette problématique nécessite une formalisation des éléments de localisation relatifs, ici sous la forme d'un modèle en triplets, utilisés comme base de référencement et un questionnement sur la prise en compte de l'imperfection du langage naturel dans le processus de modélisation. Nous proposons de faire appel aux objets spatiaux flous et nous comparerons leurs implémentations afin d'identifier l'approche la plus adaptée au contexte du secours en montagne.

**ABSTRACT.** This paper proposes a methodology for transforming a relative position description expressed in an indirect repository into an absolute position in a direct repository. This problem requires the formalization of relative location elements, here in the form of a triplet model, which is used as a basis for the transformation from relative to absolute position and the questioning on how the imperfection of the natural language in the modeling process can be taken into account. We propose to use fuzzy spatial objects and we will compare their implementations to identify the most appropriate approach to our context.

**MOTS-CLÉS :** Référentiel spatial indirect, logique floue, objets spatiaux vagues, relation spatiale étendue, secours en montagne

**KEYWORDS:** Indirect spatial referential, fuzzy logic, vague spatial objects, extended spatial relation, mountain rescue

---

## 1. Introduction

Lors de leurs interventions, les Pelotons de Gendarmerie de Haute Montagne (PGHM) peuvent être contraints à identifier manuellement la position d'une victime en recoupant des informations obtenues par téléphone (*e.g.* itinéraire emprunté, description verbale de position) avec des données métier, d'origines et de formats multiples (*e.g.* cartes numériques, guides papier). La complexité de cet exercice en fait le principal goulot d'étranglement de la phase de préparation des secours, conduisant généralement à la perte d'un temps précieux. L'ANR CHOUCAS<sup>1</sup> vise à répondre à un besoin exprimé par le PGHM de Grenoble, en proposant des méthodes facilitant le processus d'identification de la position des victimes. Cela passe notamment par des travaux sur la constitution et l'enrichissement de données provenant de sources hétérogènes, ainsi que sur le raisonnement spatial et la visualisation interactive pour l'aide à la décision (Olteanu-Raimond *et al.*, 2017).

Au sein de ce projet, notre travail part du constat que la victime décrit systématiquement sa position au moyen de relations spatiales et d'objets géographiques servant de référence, *i.e.* à l'aide d'un *référentiel spatial indirect*. Nous traiterons ici de leur transformation en des positions exprimées dans un *référentiel direct* (Clementini, 2011) compte tenu des spécificités de notre contexte. Au-delà de l'application au secours en montagne, cette problématique a pour enjeu un accès facilité à l'information géographique, plus proche de la cognition humaine que ce que permettent les outils et méthodes classiques disponibles sur SIG, comme prôné par Egenhofer et Mark (1995) avec la *naive geography*.

Cet article présentera la question de la modélisation d'une position à partir d'un référentiel indirect (Section 2), ainsi que les principes de modélisation des *relations spatiales*, retenus (Section 3). La Section 4 abordera la nécessité d'utiliser des *objets spatiaux vagues* et présentera un état de l'art des modèles et implémentations existantes, avant d'en confronter deux d'entre-elles à un cas réel (Section 5).

## 2. Modélisation d'un élément de localisation dans un référentiel indirect

Dans cette partie, nous aborderons la question de la modélisation formelle d'un *objet spatial* à partir d'un *référentiel indirect*.

### 2.1 Hypothèses de travail

Le modèle que nous présentons est basé sur l'hypothèse que le requérant (personne contactant les secours) décrit une position à l'aide d'une ou plusieurs phrases telles que « *Je suis à côté d'un arbre et en face d'un chalet* » ou « *Le blessé est à côté de moi* ».

---

<sup>1</sup> ANR CHOUCAS : *Intégration de données hétérogènes et raisonnement spatial pour l'aide à la localisation des victimes en montagne*. ANR-16-CE23-0018

Notons que nous ne travaillons pas directement à partir de phrases mais avec une liste d'éléments de localisation, (e.g. « Je suis à côté d'un arbre » et « Je suis en face d'un chalet ») extraits du discours du requérant par le secouriste et saisi au travers d'une interface également développée au sein de l'ANR CHOUCAS (dont il ne sera pas question ici).

La méthodologie que nous travaillons à définir s'articule en deux phases complémentaires, la première visant à identifier l'ensemble des positions correspondant à chaque élément de localisation donné par le requérant et la seconde constituant en l'agrégation de ces positions en vue d'identifier les positions correspondant à l'ensemble des éléments de localisation. L'imprécision est prise en compte durant tout le processus, ce n'est cependant pas le cas l'incertitude, qui n'est modélisée qu'à partir de la seconde étape, la même incertitude caractérisant alors toutes les positions décrites par un même élément de localisation. Dans ce papier, nous nous concentrons sur la première phase et ne nous intéresserons donc qu'à la modélisation de l'imprécision des éléments de localisation.

## 2.2 Formalisation des éléments de localisation

Les éléments de localisation présentent une structure commune, composée d'un sujet dont on décrit la localisation (« je », « le blessé »), d'un point de repère défini comme l'objet de référence (« un arbre », « un chalet », « moi ») et une relation indiquant la position du sujet par rapport à l'objet de référence, que nous nommerons la relation spatiale (« à côté », « en face »). De manière analogue au modèle de Vasardini *et al.* (2013), nous proposons de modéliser ces éléments de localisation par un triplet composé de ces trois éléments. Ainsi les exemples précédents peuvent être modélisés par les triplets suivants : (je, à côté, arbre), (je, en face, rivière) ; (le blessé, à côté, moi).

Cependant une phrase telle que : « je suis presque en dessous d'un télésiège » n'est pas efficacement modélisée à l'aide d'une relation spatiale, la nuance apportée par l'adverbe *presque* étant ignorée. De manière analogue au modèle des entités nommées étendues (« expanded named entity ») de Moncla *et al.* (2015), nous proposons de définir une relation spatiale étendue englobant la relation spatiale, ses modificateurs (e.g. *presque*, *complètement*, [à] 50 m) mais également le verbe, ce dernier nuancant le propos (e.g. *aller* vs *passer* ou *traverser*). La composition de la relation spatiale étendue peut varier en fonction du contexte, mais elle est toujours composée d'au moins un verbe et d'une relation spatiale, les modificateurs étant facultatifs et illimités. Un élément de localisation peut alors être modélisé avec un triplet *E* de la forme :

$$E = (S, RSE, O) \text{ avec } RSE = (V, mod, R) \quad (1)$$

Avec *S* le sujet, *RSE* une relation spatiale étendue, *O* l'objet de référence, *V* le verbe, *R* une relation spatiale et *mod*, le ou les modificateurs.

### 3. Principes de modélisation des relations spatiales étendues

Un des principaux verrous à la construction d'objets spatiaux à partir d'un référentiel indirect est l'interprétation des éléments de localisation. Bien que notre travail soit facilité par l'action du secouriste, nous restons confrontés à l'implicite et à la polysémie du langage humain. Cela pose deux principaux problèmes que nous proposons de résoudre à l'aide de deux principes de modélisation des *relations spatiales étendues*, la sécabilité et la non-bivalence.

#### 3.1 Décomposition des relations spatiales étendues

Une même relation spatiale (et par extension une relation spatiale étendue), peut avoir des significations différentes en fonction de la nature de l'objet de référence (Borillio, 1998, pp. 16-17). Prenons l'exemple de la relation spatiale *sous*. Bien que partageant une sémantique proche, son sens dans les phrases « *je suis sous un arbre* » et « *je suis sous une route* », diffère. La notion de recouvrement présente dans la première phrase est absente de la seconde. Pustejovsky (2017) parle de « *projective relations* » dans le second cas. Cette différence sémantique entre deux occurrences d'une même relation spatiale s'explique par la nature des objets de référence. Une personne peut être recouverte par un arbre et non par une route (à l'inverse d'un pont). Cela impose de prendre en compte les différentes sémantiques d'une même relation spatiale et ce en fonction des types d'objets de référence et des sujets utilisables dans ce contexte. Ces deux relations spatiales partagent toutefois une base sémantique commune : leur sujet est situé à une altitude inférieure à celle de l'objet de référence. Ce sont ces régularités sémantiques que nous proposons d'exploiter.

Notons  $Pa_o$  l'ensemble des positions situées à une altitude inférieure à un objet de référence de type  $o$  :

$$Pa_o = \cup(Pa_i | i \in o) \quad (2)$$

Avec  $o$  l'ensemble des objets spatiaux d'un même type,  $i$  une instance de  $o$  et  $Pa_i$  l'ensemble des positions situés à une altitude inférieure à celle de  $i$ .

On peut alors construire les ensembles  $Pa_r$  et  $Pa_{ab}$  contenant les positions situées respectivement à une altitude inférieure à au moins une instance d'une route ou d'un arbre. Cependant, toutes les positions appartenant à  $Pa_r$  ou  $Pa_{ab}$  ne valident pas nécessairement les éléments de localisation « *je suis sous une route* » ou « *je suis sous un arbre* ». Ces ensembles contiennent des positions situées à une grande distance de l'objet de référence qu'il serait absurde de considérer comme situées *sous* l'un de ces objets de référence.

Dès lors, il apparaît nécessaire de circonscrire la distance à l'objet de référence. Cependant (contrairement aux ensembles  $Pa_r$  et  $Pa_{ab}$ ) il n'est pas pertinent de fixer un même seuil de distance pour des objets de nature différente. La distance à partir de laquelle on ne se situe plus sous un arbre étant intuitivement plus faible que celle



à partir de laquelle on ne se situe plus sous une route. Notons  $Pb_o$  l'ensemble des positions proches d'un objet de référence de type  $o$ . Nous pouvons alors construire une approximation des zones  $Pc_r$  et  $Pc_{ab}$ , respectivement définies comme l'intersection des ensembles  $Pa_r$  et  $Pb_r$  ou  $Pa_{ab}$  et  $Pb_{ab}$ .

Nous proposons donc de prendre en compte la polysémie des relations spatiales en les modélisant comme une composition de relations spatiales à la sémantique invariable et pondérées contextuellement, que nous nommons les *relations spatiales atomiques*. Nous supposons que ce travail de décomposition peut s'appliquer à toute relation spatiale non atomique. À titre d'exemple, la relation « entouré de » peut-être modélisée comme la conjonction des relations « à gauche », « à droite », « devant » et « derrière » (Vanegas *et al.*, 2011). La principale difficulté de cette approche est qu'elle nécessite au préalable d'identifier les relations spatiales atomiques, ce qui impose une analyse manuelle des relations spatiales utilisées dans notre contexte.

### 3.2 Non-bivalence des relations spatiales étendues

Le principe de la non-bivalence découle de la nature imprécise du langage et de la difficulté de construire des objets spatiaux à partir de relations spatiales telles que « à côté » ou « sous ». La construction des ensembles  $Pc_r$  et  $Pc_{ab}$  (Section 3.1) nécessite la définition de seuils : un premier seuil représentant la différence d'altitude à partir de laquelle une personne n'est plus située sous un objet de référence, et deux seuils de distance : le premier représentant la distance à partir de laquelle on considère qu'une *personne n'est plus sous un arbre*, le second la distance à partir de laquelle on considère qu'une *personne n'est plus sous une route*.

Fixer précisément ces seuils est une opération délicate, car il est nécessaire de prendre en compte la nature de l'objet de référence (route vs arbre), mais aussi de son type (route vs chemin) et l'inconstance des locuteurs, liée à une expression « naïve » des localisations (Egenhofer et Mark, 1995). Le choix d'un seuil trop restrictif ou trop lâche sera nécessairement source d'un grand nombre de faux, positifs ou négatifs.

Pour pallier ce problème, la solution la plus rependue dans la littérature consiste à construire des objets *spatiaux vagues*, *i.e.* des objets aux limites nettes, capables de prendre en compte l'imprécision. Ce type d'objet, présenté ci-dessous, permet de délimiter des zones de façon non bivalente.

## 4. Modélisation des objets spatiaux vagues

Cette partie présentera un aperçu des différents modèles et implémentations proposés dans la littérature pour modéliser des *objets spatiaux vagues* (Dilo, 2006). Nous en présenterons ensuite une sélection, au regard de notre problématique.

#### 4.1 État de l'art

Le recours à la modélisation d'objets vagues naît de la nécessité de modéliser des objets spatiaux dont les caractéristiques rendent caduque la modélisation à l'aide d'objets spatiaux aux frontières nettes. C'est notamment le cas lorsque l'on cherche à modéliser des objets dont la frontière ne peut pas être définie de manière précise. Citons l'exemple de la rivière polluée proposé par Dilo (2006). Dans ce cas, il est difficile de définir la zone de la rivière polluée puisque le degré de pollution décroît graduellement à partir de la source. Fixer la frontière d'un tel objet conduira à un choix discutable, puisque dépendant d'un seuil fixé empiriquement.

De nombreuses approches et implémentations des objets spatiaux vagues ont été proposées dans la littérature. Clementini (2008) distingue trois catégories : les modèles définissant des objets aux frontières étendues (« *objects with broad boundaries* »), les modèles basés sur la théorie des ensembles flous (Zadeh, 1965) et ceux basés sur la théorie des probabilités.

La catégorie des modèles aux frontières étendues regroupe les modélisations d'objets spatiaux remplaçant les traditionnelles frontières nettes par des frontières non nettes. Le modèle *egg-yolk* proposé par Cohn et Gotts (1996) est probablement le plus connu dans cette catégorie. On peut également citer le modèle min-max (Bejahoui *et al*, 2009) qui est une amélioration du modèle *egg-yolk* ou le modèle de Clementini (2008). Selon la définition proposée par Clementini (2008) un objet aux frontières vagues est composé de deux régions  $ra$  et  $rb$  avec  $ra \subseteq rb$ .  $\partial ra$  correspond à la frontière intérieure et  $\partial rb$  à la frontière extérieure de l'objet spatial. La frontière de l'objet spatial est alors définie comme la zone comprise entre la frontière intérieure et la frontière extérieure.

Les modèles basés sur la théorie des ensembles flous (voir la Figure 1), tel que celui proposé par Schneider (1999), proposent de faire une analogie entre un objet spatial et un ensemble flou. Zadeh (1969) définit un ensemble flou  $A$  comme un couple composé d'un ensemble  $X$  et d'une fonction  $f_A$  nommée fonction d'appartenance. La fonction  $f_A$ , associe à chaque élément de  $X$  une valeur comprise dans l'intervalle  $[0,1]$ , nommée degré d'appartenance :

$$A = (X, f_A), \text{ avec } f_A: X \rightarrow [0,1] \quad (3)$$

L'ensemble des éléments ayant un degré d'appartenance non nul permet de définir le *support*  $S$  de l'ensemble (cf. eq. 5). Le *noyau*  $N$  de l'ensemble correspond, quant à lui, à l'ensemble des éléments ayant un degré d'appartenance égal à 1. Le noyau,  $N$ , et le support,  $S$ , de l'ensemble flou  $A$  sont des ensembles nets.

$$N(A) = \{x \in X | f_A(x) = 1\} \quad (4)$$

$$S(A) = \{x \in X | f_A(x) > 0\} \quad (5)$$

Le noyau et le support de l'ensemble  $A$  sont des cas particuliers de coupes-alpha (*alpha-cuts*) définies comme :

$$A^{>\alpha} = \{x \in X | f_A(x) > \alpha\} \quad (6)$$

Schneider (1999) définit un équivalent flou à tous les objets spatiaux habituels. Ainsi, les points, les lignes et les polygones sont respectivement déclinés en points, lignes et polygones flous. Un point flou  $pf$  est défini comme un point ayant un degré d'appartenance non nul. Par extension une ligne et une région floue  $lf$  et  $rf$  sont définies comme des ensembles particuliers de points flous avec des contraintes spécifiques. Dilo *et al.* (2007) ont proposé une implémentation de ce modèle. Un point flou est implémenté comme un triplet contenant les coordonnées du point et son degré d'appartenance. De manière analogue au modèle théorique, une ligne floue est une liste ordonnée de points flous. Les auteurs proposent de reconstruire une approximation de la ligne floue à l'aide d'une interpolation linéaire entre chaque couple de points. La même démarche est proposée pour l'implémentation des régions floues. Ainsi, les limites du noyau et du support de la région floue sont stockées sous la forme de polygones nettes et pour chaque point de la région floue une approximation du degré d'appartenance est calculée par interpolation.

Vanegas *et al.* (2011) proposent également de modéliser des objets spatiaux flous, mais avec une implémentation raster. L'espace traité est divisé en cellules de même taille pour lesquelles est calculé un degré d'appartenance. Cette approche permet de construire en extension l'ensemble des positions validant les règles de construction définies. Cette approche présente l'avantage de faciliter les opérations inter-objets, mais sa précision est fortement dépendante de la résolution choisie.

Une autre proposition, basée sur la théorie des ensembles flous, est d'enregistrer une approximation des objets spatiaux flous à l'aide de coupes-alpha (de Runz, 2008 ; Zoghliani *et al.*, 2016). Cette implémentation offre la possibilité de stocker des objets spatiaux flous sans recourir à des développements spécifiques, contrairement à l'approche de Dilo *et al.* (2007), mais les opérations inter-ensembles (union, intersection, etc.), demandent toutefois des développements *ad hoc*, contrairement à l'approche raster, en contrepartie plus volumineuse.

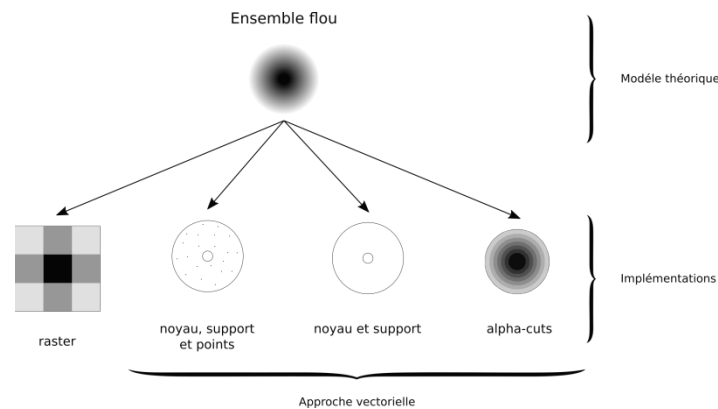


Figure 1. Implémentations d'objets spatiaux flous

Enfin, Tøssebro et Nygård (2002) proposent une troisième catégorie de modèle, basée sur la théorie des probabilités. De manière analogue au travail de Schneider (1999), ils proposent d'étendre les concepts de points, lignes et régions en leurs équivalents incertains. Dans ce cas, les points, lignes et régions sont inclus au sein d'une région nette, définissant l'ensemble des positions possibles. L'incertitude est modélisée à l'aide d'une fonction de masse, attribuant à chaque point de la région nette une probabilité d'accueillir l'objet spatial incertain.

#### **4.2 Comparaison et choix d'un modèle et de son implémentation**

Le choix de notre approche est contraint par le besoin d'agréger l'information issue de la modélisation de plusieurs éléments de localisation, mais aussi de plusieurs relations spatiales atomiques. Tous les modèles présentés permettent ce type de manipulation, mais tous n'offrent pas la même simplicité.

Les modèles discrets, tels que *min-max* ou *egg-yölk*, ont pour principal défaut leur première qualité, à savoir la simplicité. Ils ne permettent pas de prendre en compte la variation d'appartenance entre leurs deux frontières. Le calcul du degré d'appartenance d'une position à une région floue n'est donc envisageable qu'au prix d'une hypothèse sur la variation du degré d'appartenance entre les frontières. Les modèles proposés par Schneider (1999), Tøssebro et Nygård (2002) permettent, quant à eux, de modéliser une variation quelconque du degré d'appartenance. Bien entendu, les implémentations de ces modèles échantillonnent cette valeur, mais la précision de cette dernière, notamment dans les cas de variation non linéaire du degré d'appartenance, reste plus importante que celle des modèles *min-max* ou *egg-yölk*.

Parmi les modèles permettant de modéliser des variations non linéaires, nous avons choisi de travailler avec un modèle basé sur la théorie des ensembles flous, ces modèles étant les plus adaptés à nos contraintes. Toutefois, il est nécessaire de choisir une implémentation parmi celles présentées précédemment. Nous avons décidé de ne pas travailler à partir du modèle présenté par Dilo (2007), bien qu'il s'agisse d'une approche intéressante, puisque combinant à la fois les qualités d'une approche vectorielle (définition des frontières) et matricielle (meilleure précision de l'échantillonnage du degré d'appartenance). Cependant, cette implémentation est peu efficace pour manipuler des opérations inter-ensembles, fortement utilisées dans notre cas. Pour trancher entre une implémentation sous la forme d'*alpha-cuts* (de Runz, 2008), solution la plus économe en mémoire et une implémentation raster (Vanegas, 2011), plus précise, nous avons choisi de confronter ces deux approches sur un cas particulier, celui de l'élément de localisation : « *je suis sous une ligne électrique* ».

### **5. Cas d'application**

En vue d'une sélection nous allons présenter une comparaison de deux implémentations des objets spatiaux flous présentés ci-dessus, à savoir les *alpha-cuts* (de Runz, 2008) et l'approche raster (Vanegas, 2011). Ces deux

implémentations seront appliquées au même cas, issu d'une situation réelle. La victime a indiqué qu'elle était située sous une ligne à haute-tension, élément de localisation correspondant au triplet (*je, (suis, sous), une ligne à haute tension*).

### 5.1 Identification des relations spatiales atomiques et définition des fonctions d'appartenance

Comme illustré dans la Section 3.1, être *sous* un objet, quelle que soit sa nature, implique que le sujet soit proche et situé à une altitude inférieure à l'objet auquel il se réfère. Dans notre contexte, caractérisé par des dénivelés pouvant être importants, ces deux contraintes sont fortement indépendantes. Nous définissons donc deux règles : 1) le sujet doit être proche de l'objet de référence et 2) le sujet doit avoir une altitude inférieure à l'objet de référence. Il est nécessaire de fixer des seuils pour transformer ces descriptions qualitatives en règles exploitables, ici nous nous contenterons de fixer des seuils génériques, laissant le traitement de leurs variabilités locales à une exploration ultérieure. Le recours aux objets spatiaux flous permet de prendre en compte l'imprécision inhérente à ce type de transformations, il est toutefois nécessaire de fixer manuellement deux seuils :  $d_S$  représentant la distance à partir de laquelle une position n'est plus *proche* d'une ligne à haute-tension (limite du support) et  $d_N$ , la distance en-deca de laquelle une position est *proche* d'une ligne à haute-tension (limite du noyau). La Figure 2 présente ces seuils et la forme de la fonction d'appartenance qui en découle. L'ensemble des positions proches d'une ligne à haute tension est défini par la fonction d'appartenance suivante :

$$f_1(d) = \begin{cases} 1 & \text{si } d < d_N \\ 0 & \text{si } d > d_S \\ 1 - \frac{d-d_N}{d_S-d_N} & \text{sinon} \end{cases} \quad (7)$$

Où  $d$  représente la distance à l'axe médian de la ligne à haute tension. Pour cette application, en collaboration avec le PGHM de Grenoble, les seuils retenus sont de 50 m pour  $d_N$ , et de 100 m pour  $d_S$ .

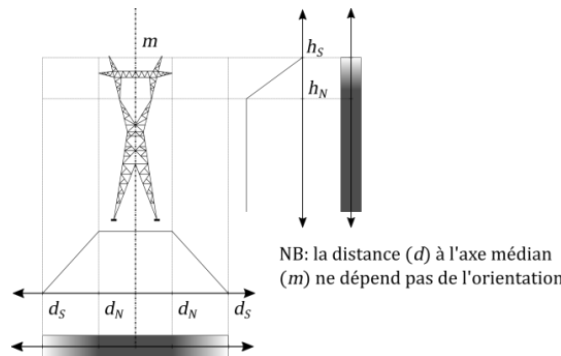


Figure 2. Définition de l'ensemble flou « je suis sous une ligne électrique »

De la même manière, l'ensemble contenant toutes les positions situées à une altitude inférieure à celle d'une ligne à haute tension est défini à travers la fonction d'appartenance suivante :

$$f_2(H, h) = \begin{cases} 1 & \text{si } H - h > h_N \\ 0 & \text{si } H - h < h_S \\ \frac{(H-h)-h_N}{h_S-h_N} & \text{sinon} \end{cases} \quad (8)$$

Avec  $H$  la hauteur de la ligne à haute-tension la plus proche et  $h$  l'altitude de la position,  $h_N$  et  $h_S$  la différence entre  $H$  et la hauteur, respectivement, du noyau et du support. Les seuils ont été fixés empiriquement à 5 m pour  $h_N$  et à 0 m pour  $h_S$  (soit le sommet du pylône).

Enfin, l'ensemble contenant toutes les positions situées sous une ligne à haute tension est défini à partir des fonctions d'appartenances décrites par les équations 7 et 8 de la manière suivante :

$$f_3(d, H, h) = \min(f_1(d), f_2(H, h)) \quad (9)$$

Où  $\min$  représente l'opérateur d'intersection proposé par Zadeh (1965), est défini comme le minimum des deux valeurs d'appartenance.

## 5.2 Application

Les deux implémentations présentées, portent sur la région de Bourg d'Oisans, au sud-est de Grenoble. Deux sources de données ont été utilisées. Les lignes à haute tension, ainsi que leurs pylônes proviennent de la BD TOPO produite par l'IGN. Le MNT, d'une résolution de cinq mètres, est quant à lui extrait de la BD ALTI de 2017.

La première implémentation testée repose sur la méthodologie proposée par Vanegas *et al.* (2011). L'ensemble des positions étudiées, délimité par une aire d'étude précédemment définie, est échantillonné par une grille à la résolution calquée sur le MNT. Nous calculons alors la valeur du degré d'appartenance pour chacune de ces cellules, permettant ainsi de définir un ensemble flou en extension. Dans notre cas, nous construisons trois rasters (Figure 3), représentant autant d'ensembles flous, comme défini dans les équations 7, 8 et 9. Nous avons choisi de rasteriser la couche *ligne électrique* de la BD TOPO, contenant l'ensemble des portions aériennes de lignes à haute et très haute tension. À l'aide de cette couche intermédiaire nous mesurons la distance planimétrique séparant chaque cellule du raster de la cellule la plus proche occupée par une ligne électrique. Cette mesure de distance est utilisée comme paramètre de la fonction d'appartenance présentée en Eq. 7 et dont le résultat est présenté par la Figure 3(a).

Le calcul du degré d'appartenance à l'ensemble des positions situées à une altitude inférieure à celle d'une ligne à haute-tension nécessite, quant à lui, de calculer la différence entre l'altitude de la position et celle des lignes à haute tension. Il s'est avéré nécessaire de calculer une estimation de la hauteur de la ligne

à haute tension en toute cellule traversée par une ligne à haute tension. Nous avons estimé cette valeur à l'aide d'une interpolation linéaire limitée à deux points (les deux pylônes les plus proches de la cellule), avant de rogner ce résultat à l'aide de l'empreinte au sol des lignes à haute tension précédemment calculées. Cette nouvelle donnée, enrichissant la rasterisation des lignes à haute-tension a été utilisée comme paramètre de la fonction d'appartenance présentée en Eq. 8. La Figure 3(b) représente le degré d'appartenance ainsi calculé.

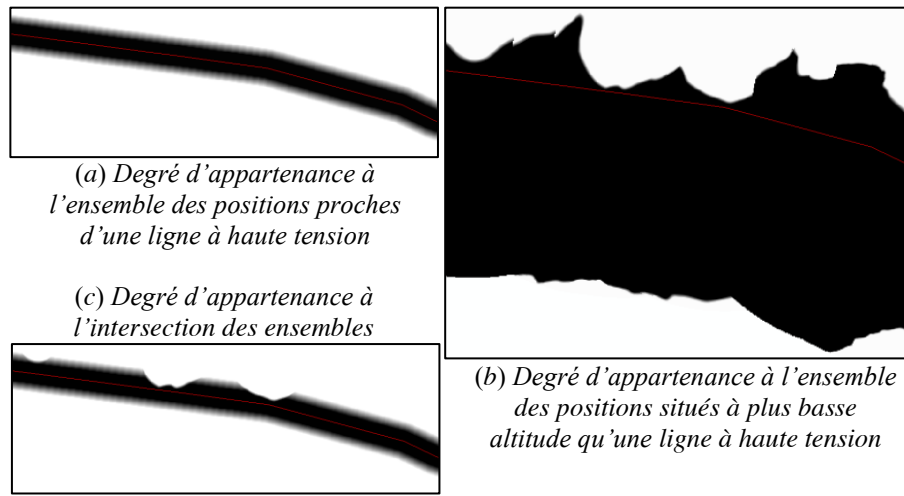


Figure 3. Modélisation d'objets spatiaux flous : approche raster

Les deux ensembles flous (Figures. 3(a) et 3(b) permettent de construire l'ensemble représentant les positions situées *sous une ligne électrique*. La fonction d'appartenance de ce dernier prend trois paramètres (cf. Eq. 9). Nous remarquons que la zone construite est sensiblement différente de celle que nous aurions obtenue si nous n'avions pris en compte la seule distance (Figure 3(a)). Certaines positions, très proches d'une ligne à haute-tension peuvent, en effet, être situées à des altitudes dépassant fortement le sommet des pylônes avoisinants, démontrant l'intérêt de prendre en compte la différence d'altitude.

La seconde implémentation testée est basée sur l'approche utilisée par de Runz (2008) et Zoghلامي et al. (2016). Nous effectuons ici tous nos calculs à l'aide d'une approche vectorielle. Cependant, aucun système d'information géographique ou plus généralement logiciel permettant la manipulation de données spatiales vectorielles ne permet, du moins à notre connaissance, de manipuler des géométries

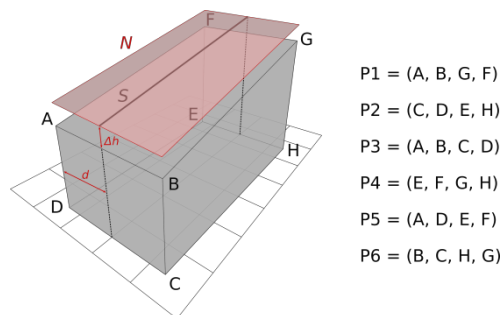


Figure 4. Plans utilisés pour définir une alpha-cut

tridimensionnelles telles que celles que nous souhaitons construire. Nous nous sommes donc orientés vers les outils proposés par la bibliothèque CGAL, spécialisée dans la géométrie algorithmique.

La démarche suivie a consisté à créer des polyèdres, représentant les alpha-cuts tridimensionnelles de l'ensemble flou contenant les positions sous une ligne électrique. Comme illustré par la Figure 4, pour chaque segment  $S$  d'une ligne à haute tension, nous définissons six plans, délimitant autant de demi-espaces et dont l'intersection permet de définir un volume fini. Le premier plan (P1) est parallèle au plan  $N$  incluant le segment  $S$  et une droite horizontale perpendiculaire à  $S$ . La différence d'altitude ( $\Delta h$ ) entre les plans P1 et  $N$  varie en fonction de l'alpha-cut modélisée. Le second plan (P2) représente la limite inférieure du polyèdre, son  $z$  est nul et constant. Les plans P3 et P4 représentent l'avant et l'arrière du polyèdre (par rapport à la direction de la ligne électrique). Ils sont perpendiculaires au plan P2 et passent respectivement par le premier et dernier point du segment  $S$ . Enfin, les plans P5 et P6 représentent la gauche et la droite du polyèdre. Ils sont verticaux et parallèles au segment  $S$ . Comme pour le plan P1, leur distance ( $d$ ) au segment  $S$  varie en fonction de l'isovolume modélisé.

Pour modéliser les alpha-cuts, nous faisons varier la hauteur du plan P2 et l'éloignement des plans P5 et P6 par rapport au segment  $S$ . En identifiant les valeurs de hauteur et d'éloignement correspondant au degré d'appartenance défini on peut construire les *alpha-cuts* désirées. La Figure 5(a) montre (sans prise en compte du MNT) trois *alpha-cuts* (0 ; 0,5 et 1) construites à partir d'un segment de ligne à haute-tension. En intersectant ces géométries avec le MNT (Figure 5(b) nous

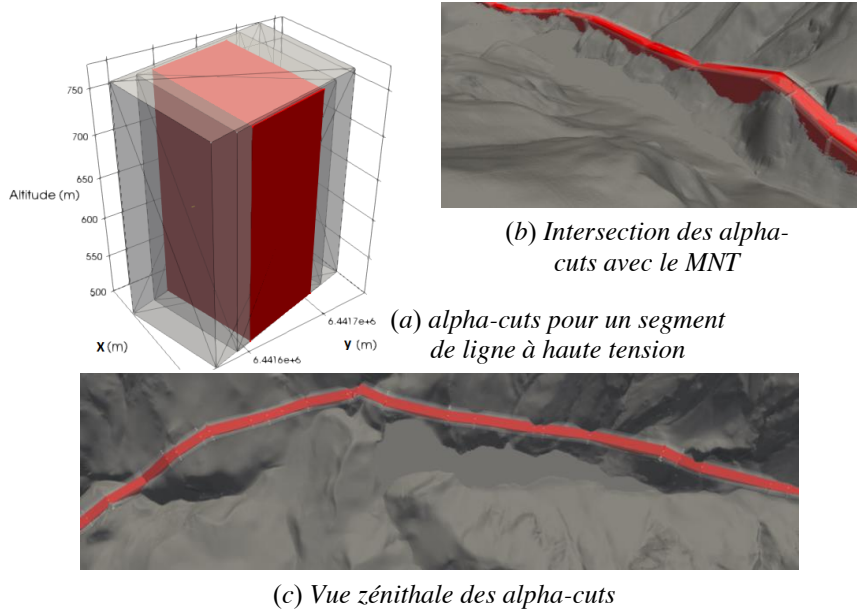


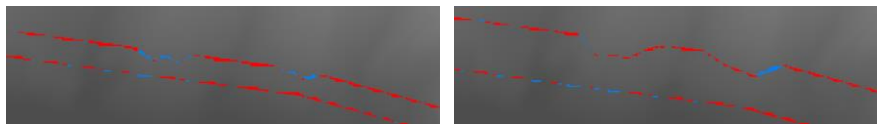
Figure 5. Modélisation d'objets spatiaux flous : approche vectorielle



pouvons projeter les *alpha-cuts* au sol et ainsi construire une carte en deux dimensions (Figure 5(c) représentant les *alpha-cuts* de l'ensemble flou des positions situées sous une ligne haute-tension dans notre aire d'étude.

### 5.3 Comparaison

Cette double implémentation d'un même élément de localisation a fait apparaître les avantages et limites des deux approches. Notre besoin de travailler en trois dimensions est source de nombreuses complications liées à la difficulté de manipulation de géométries 3D. L'implémentation en *alpha-cuts* a nécessité l'utilisation de la bibliothèque CGAL (aucun SIG n'étant envisageable) et un travail de développement non négligeable (cf. tableau 1), pour un résultat comparable à celui de l'approche raster (cf. figure 6). L'approche par *alpha-cuts* a été totalement automatisée, alors que notre implémentation de l'approche raster nécessite encore plusieurs opérations manuelles pour enchaîner les étapes (cf. tableau 1, la durée des opérations manuelles n'est pas comptabilisée dans le temps d'exécution de l'approche raster).



*Figure 6 Comparaison des zones construites à l'aide des approches raster et alpha-cuts. En rouge figurent les pixels uniquement présents dans l'estimation par raster et en bleu les zones uniquement présentes dans l'estimation par alpha-cuts. L'image de gauche représente la comparaison des deux noyaux et celle de droite la comparaison des deux supports*

La figure 6 illustre les différences d'estimation de la zone « sous une ligne électrique » entre l'approche raster et l'approche par *alpha-cuts*. Ici les différences, représentées sous la forme d'un raster, dont la maille correspond à celle du MNT, sont de l'ordre du pixel. Les deux méthodes aboutissent à des résultats similaires. On peut cependant remarquer que l'approche par raster tend, comparativement à l'approche par *alpha-cuts*, à surévaluer légèrement la zone correspondant à l'élément de localisation modélisé. Mais cette comparaison illustre avant tout les artéfacts nés d'une implémentation imparfaite. L'approche raster offre de meilleurs résultats à proximité de la jointure des segments de ligne, notre implémentation de l'approche par *alpha-cuts* n'opérant pas de traitement particulier en vue de modéliser ces jointures. Inversement l'approche par *alpha-cuts* est plus efficace lorsque l'estimation de la hauteur de la ligne à haute-tension utilisée par l'approche raster est trop grossière, ce qui est notamment le cas lorsque plusieurs lignes sont à proximité. Il s'agit toutefois de problèmes d'implémentation pouvant être corrigés dans les deux cas, même si la quantité de travail supposée nécessaire est plus importante dans le cas de l'approche par *alpha-cuts*.

Dans le cas où les deux approches ne produisent pas d'artéfacts (comme c'est le cas pour la zone représentée par la figure 6), les deux estimations sont extrêmement

proches, ce qui nous pousse naturellement à privilégier l'approche présentant le meilleur ratio qualité de l'estimation / difficulté de mise en place. Il nous semble donc plus pertinent de travailler, comme le propose Vanegas *et al.* (2011), à l'aide de rasters. Les règles utilisées pour définir l'ensemble flou à l'aide de l'approche raster présentent l'avantage d'être simples et peu nombreuses, ce qui laisse également supposer une meilleure capacité de généralisation de la méthodologie.

*Tableau 1 Comparaison des temps de développement, d'exécution et estimation des temps d'amélioration des deux approches présentées.*

	Raster	Alpha-cuts
Temps de développement	~ <b>5 h</b>	~ 60 h
Temps d'exécution	~ <b>2 min</b>	~ 5 min
Estimation de l'importance des développements à venir	+	+++

## 6. Conclusion

Dans cet article, nous avons présenté notre démarche visant à construire les zones de présence possible d'un objet dont la position est décrite à partir d'un référentiel spatial indirect à l'aide d'éléments de localisation. Ces derniers ont été modélisés sous la forme de triplets, de manière analogue aux modèles présentés dans la littérature. Nous avons cependant étendu ce modèle en y ajoutant le concept de *relations spatiales étendues*, apportant plus de précision. À partir de ce modèle, nous avons présenté le concept de décomposition des relations spatiales, visant à exploiter les redondances sémantiques des relations spatiales. Nous avons également présenté, les problématiques liées à l'incertitude et l'intérêt d'adopter une démarche de modélisation des relations spatiales non-bivalentes, à l'aide de la théorie des ensembles flous. Pour finir, nous avons proposé une application comparative de deux implémentations de la théorie des ensembles flous, démontrant l'adéquation de l'approche raster à notre problématique.

Nous souhaitons désormais travailler à l'extension du modèle en triplets pour y adjoindre la gestion des descriptions de mouvement (« *j'ai longé la route* »), la prise en compte de la dimension temporelle (« *j'étais* » vs « *je suis* ») ou les référencements chaînés (« *Je suis près d'un pont qui enjambe la rivière* »).

La décomposition des relations spatiales nécessite, quant à elle, un important travail d'analyse, visant à développer une méthodologie d'identification des redondances sémantiques, sur la base d'un recensement des relations spatiales les plus présentes dans notre contexte.

Une autre tâche importante est d'automatiser la définition des seuils. Dans l'exemple présenté ces derniers sont définis empiriquement. Cependant, cette approche pose deux problèmes, le premier étant que les seuils sélectionnés ne sont pertinents que pour un type d'objet de référence donné, le second est qu'il serait

intéressant de faire varier ces seuils localement pour prendre en compte certaines spécificités du terrain. A titre d'exemple, l'élément de localisation « *je suis sous une ligne électrique* » nécessite de prendre en compte tout type de ligne et non plus seulement les lignes à haute-tension, par conséquent il sera nécessaire de faire varier les seuils précédemment présentés en fonction du type de la ligne électrique (dont dépendent sa hauteur et sa largeur). Le second aspect est que dans un contexte montagneux, la hauteur d'une ligne à haute tension par rapport au sol peut varier fortement, au gré des ruptures de pente. On peut postuler qu'il en découle une forte variabilité dans la perception du recouvrement, modélisable par une variation locale des seuils  $d_s$  et  $d_N$  (cf. eq. 7). Une des pistes envisagée est de recourir à des méthodes d'apprentissage, dont l'application à l'identification de relations spatiales floues a déjà été expérimentée Xu et al. (2006). Bien que cette approche soit séduisante, la nécessaire constitution d'une base d'apprentissage est un défi de taille compte-tenu de l'importance des spécificités locales.

Nous souhaitons également aborder la question de l'insertion de la méthodologie présentée au sein du processus de localisation de la victime. Ici nous avons choisi de calculer directement la zone, mais il serait envisageable d'opter pour une approche itérative, fondée sur des raffinements successifs de la zone estimée. Cette solution, qui peut apporter un gain de performances non négligeable, pose toutefois de nouvelles questions méthodologiques et nous expose au risque de voir apparaître des faux négatifs liés au MAUP (Openshaw, 1984).

A plus long terme, nous souhaitons intégrer la prise en compte de l'incertitude à la méthodologie développée. Comme nous l'avons précédemment expliqué, nous envisageons de traiter l'incertitude (e.g. « *je crois que je suis sous une ligne électrique* ») lors de l'agrégation des zones correspondant aux éléments de localisation, en vue de construire la zone de présence finale de la victime. La prise en compte de l'incertitude nécessite de coupler la théorie des sous-ensembles flou à une théorie permettant de modéliser l'incertitude telle que la théorie des possibilités ou celle des fonctions de croyances.

## Bibliographie

- Bejaoui L., Pinet Fr., Bedard Y., Schneider M. (2009). Qualified Topological Relation Between Spatial Objects with Possible Vague Shape. *International Journal of Geographical Information Science*, vol. 23, n° 7, p. 877-921.
- Borillo A. (2009). *L'espace et son expression en français*, Ophrys, Paris.
- Clementini E. (2008). Objects with Broad Boundaries. *Encyclopedia of GIS*. New-York, Springer, p. 793-798.
- Cohn A.G., Gotts N.M. (1996). The "egg-yolk" representation of regions with intermediate boundary. *Geographical Objects With Intermediate Boundary*. London, Taylor & Francis, p. 171-187.
- De Runz C. (2008). Fouilles archéologiques: à la recherche d'éléments représentatifs. *Atelier Fouille de données complexes*, Sofia Antipolis.

- Dilo A. (2006). *Representation of and reasoning with vagueness in spatial information. A system for handling vague objects*. Wageningen University.
- Dilo A., de By R., Stein A. (2007). A System of Types and Operators for Handling Vague Spatial Objects. *International Journal of Geographical Information Science*, vol. 21, n° 4, p. 397-426.
- Egenhofer M., Mark D. (1995). Naive Geography. *International Conference on Spatial Information Theory*, Springer, Berlin.
- Moncla L., Gaio M. (2015). A Multi-Layer Markup Language for Geospatial Semantic. *9th Workshop on Geographic Information Retrieval 2015*, ACM, Paris.
- Olteanu-Raimond A.-M., Davoine P.-A., Gaio M., Gouardères E., Van Damme M.-D., Villanova-Olivier M., Brasebin M., Domingues C., Duchêne C., Favre O., Mustière S., Devin Fl., le Nir Y., Moncla L., Bouveret S., Genoud Ph., Gensel J., Ziebelin D. (2017). Projet CHOUCAS : Intégration de données hétérogènes et raisonnement spatial pour l'aide à la localisation des victimes en montagne. *Spatial Analysis and GEomatics 2017*, SAGEO 2017, Rouen.
- Openshaw, S., 1984. The Modifiable Areal Unit Problem. Pustjovsky J. (2017). ISO-Space: Annotating Static and Dynamic Spatial Information. *Handbook of Linguistic Annotation*. Netherlands, Springer, p. 989-1024.
- Schneider M. (2008). Fuzzy Spatial Data Types for Spatial Uncertainty Management in Databases. *Handbook for Research on Fuzzy Information Processing in Databases*. Hershey, Information Science Reference, p. 490-515.
- Schneider M. (1999). Uncertainty Management for spatial data in databases: fuzzy spatial data types. *Advances in Spatial Databases*. Springer, Berlin, p. 330-351.
- Tøssebro E., Nygård M. (2002). An advanced discrete model for uncertain spatial data. *Lecture Notes in Computer Science*. Springer, Berlin, p. 37-51.
- Vanegas M.-C., Bloch I., Inglada J. (2011). A Fuzzy Definition of the Spatial Relation "Surround". *EUSFLAT Conference 2011*, Atlantis Press, Aix-les-Bains.
- Vasardani M., Timpf S., Winter S., Tomko M. (2011). From Descriptions to Depictions: A Conceptual Framework. *Spatial information theory*, Springer, Berlin.
- Xu J., Yao C. (2006). Formalizing natural-language spatial relations descriptions with fuzzy decision tree algorithm. *Geoinformatics 2006*. SPIE. Wuhan.
- Zadeh L. (1965). Fuzzy sets. *Information and Control*, vol. 8, n° 3, p. 338-353.
- Zoghلامي A., de Runz C., Herman A. (2016). F-perceptory : an Approach for Handling Fuzziness of Spatiotemporal data in Geographical Databases. *International Journal of Spatial, Temporal and Multimedia Information Systems*, vol. 1, n° 1, p. 30-62.

---

# Méthodes de validation spatiale en modélisation des changements d'occupation et d'usage des sols : le cas des modèles de potentiel de transition

Rahim Aguejdad<sup>1</sup>

1. UMR TETIS, Maison de la Télédétection  
500 rue Jean-François Breton, 34093 Montpellier, France  
[rahim.aguejdad@cnr.fr](mailto:rahim.aguejdad@cnr.fr)

---

*Abstract. Using land use and cover change (LUCC) models for the urban growth planning, environmental assessment, and decision-making needs establishing an appropriate level of confidence in their performance. The objective of this research is to explore the importance of using multiple assessment techniques in order to fairly evaluate the predictive performance of land change models. Accordingly, an application is conducted by using the Land Change Modeler for Ecological Sustainability (LCM) which is an empirical and transition potential model. LCM is applied to model the agricultural to developed areas transition in Rennes metropolitan area (France). The land demand for the prediction and reference year 2005 is estimated using the Markov Chain model; whereas, the transition potential map is implemented using the Multi-Layer Perceptron Neural Network (MLP) method based on historical changes and driving variables between 1984 and 2000. The model validation is carried out by assessing the quantity, allocation, and spatial patterns accuracy. Actually, a variety of the most commonly used validation techniques is used. Additionally, the quantity disagreement is analyzed by taking into consideration the temporal non-stationarity of change rates over the calibration and prediction intervals, the model ability to reproduce the past amount of change, and the time duration of the prediction interval. At the study area level, the correctness and disagreement analysis shows that LCM performs better at predicting the amount than the allocation of developed areas. Additionally, landscape metrics reveal that LCM tends to predict a fragmented urban form, which seems evident because of the large number of the individual urban patches. At the municipality level, the error budget analysis shows that the model performance, which varies highly between different subareas, needs to be improved. Moreover, the cross tabulation between the transition potential map and both the observed and predicted agricultural to developed areas transitions reveals that the order of the transition potential values does not perfectly fit the observed change; whereas, the predicted change is not solely limited to areas with high potential. In addition to that, the results show that selected training and validation time periods, significantly, influence the predictive performance estimation and outputs of data-driven LUCC models.*

*Keywords: Land-change, non-stationarity, modeling, data-driven models, transition potential, calibration, validation.*

*Résumé. L'usage des modèles de changement d'occupation et d'usage des sols pour la planification de la croissance urbaine, l'évaluation environnementale et l'aide à la décision, nécessite d'établir un niveau de confiance approprié dans leur performance prédictive. L'objectif de cette recherche est d'explorer l'apport de l'usage de différentes techniques de validation afin d'évaluer la performance des modèles de simulation. L'application utilisant le modèle empirique et spatialement explicite LCM porte sur la modélisation de l'artificialisation des surfaces agricoles à l'échelle de Rennes Métropole (France) dont les trajectoires d'évolution des changements d'occupation et d'usage des sols entre 1984 et 2005 ont préalablement été reconstruites. Les probabilités de transition et la quantité de changement pour l'année de prédiction 2005 sont estimées par extrapolation des tendances passées à l'aide d'un module probabiliste utilisant les chaînes de Markov. Ensuite, la carte du potentiel de transition est réalisée à l'aide d'un perceptron multicouche appartenant à la famille des réseaux de neurones et calibré grâce aux changements observés entre 1984 et 2000. La validation des sorties du modèle est réalisée en analysant les erreurs de quantité et d'allocation spatiale dues à la différence entre le changement observé et le changement simulé en 2005. En effet, plusieurs techniques de validation sont utilisées. De plus, l'erreur due à la quantité est analysée en tenant compte de la non-stationnarité des taux de changement durant les périodes de calibration et de simulation, de la durée de l'intervalle de simulation, ainsi que de la capacité du modèle à reproduire les rythmes de changement passé. A l'échelle de la zone d'étude, l'analyse des erreurs montre que le modèle est plus performant pour prédire la quantité que l'allocation des zones artificialisées. De plus, les métriques paysagères révèlent que le modèle tend à prédire une forme urbaine fragmentée en raison de la présence d'un grand nombre de fragments urbains isolés. A l'échelle communale, l'analyse des erreurs montre que la performance du modèle, qui varie fortement d'une commune à l'autre, nécessite d'être améliorée. Par ailleurs, le croisement entre la carte du potentiel de transition et les cartes de transition observée et simulée des zones agricoles vers les surfaces artificialisées révèle que l'ordre des valeurs du potentiel de transition ne correspond pas parfaitement au changement observé ; tandis que le changement simulé n'est pas limité aux seules zones à fort potentiel de transition. Les résultats montrent aussi que le choix des périodes de calibration et de validation influence significativement la performance prédictive estimée et les sorties des modèles.*

*Mots-clés : Changement d'occupation du sol, non-stationnarité, modèles empiriques, potentiel de transition, calibration, validation.*

Aguejdad, R., Houet, T., Hubert-Moy, L. (2017). Spatial validation of land-use change models using multiple assessment techniques: A case study of transition potential models. *Environmental Modeling and Assessment*, pp. 1-16.

---

## Modélisation spatiale fine de la distribution de moustiques *Aedes albopictus* par forçage météorologique et intégration de données environnementales et *in situ*

Marti Renaud<sup>1</sup>, Catry Thibault<sup>1</sup>, Marie Demarchi<sup>2</sup>, Grégory L'Ambert<sup>3</sup>, Myriam Cros<sup>3</sup>, Jean-Sébastien Dehecq<sup>4</sup>, Nadine Dessay<sup>1</sup>, Morgan Mangeas<sup>5</sup>, Emmanuel Roux<sup>1,6</sup>, Vincent Herbreteau<sup>7</sup>, Annelise Tran<sup>8,9,10</sup>

1. UMR 228 ESPACE-DEV, Institut de Recherche pour le Développement (IRD)  
500 rue Jean-François Breton 34000 Montpellier, France
2. Ingénieur indépendant, Maison de la télédétection, France
3. EID Méditerranée, Montpellier, France
4. Agence de Santé Océan Indien, Sainte-Clotilde, Ile de la Réunion
5. ESPACE-DEV, UMR228 IRD/UM/UR/UG/UA, Institut de Recherche pour le Développement (IRD), Nouvelle-Calédonie, France
6. Fundação Oswaldo Cruz (Fiocruz), Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (ICICT), Rio de Janeiro (RJ), Brasil
7. ESPACE-DEV, UMR228 IRD/UM/UR/UG/UA, Institut de Recherche pour le Développement (IRD), Institut Pasteur du Cambodge, Phnom Penh, Cambodge
8. CIRAD, UMR TETIS, Sainte-Clotilde, Ile de la Réunion
9. CIRAD, UMR ASTRE, Sainte-Clotilde, Ile de la Réunion
10. Université Montpellier, Montpellier, France

Contact : [annelise.tran@cirad.fr](mailto:annelise.tran@cirad.fr)

---

### RESUME.

Le moustique *Aedes* (*Stegomyia*) *albopictus* (Skuse) (Diptera : Culicidae) est une espèce particulièrement invasive, qui démontre de remarquables capacités d'adaptation à des conditions climatiques multiples. Le suivi de son aire de diffusion répond à une préoccupation importante de santé publique, puisque cette espèce présente, en plus d'une forte capacité de nuisance diurne, la capacité de transmettre les virus de la dengue, du chikungunya, et du Zika. Sa forte adaptation aux milieux anthropisés, et son caractère hautement invasif justifient une politique de surveillance appropriée. Dans cette étude publiée dans la revue *International Journal of Environmental Research and Public Health* sous le titre « *A rainfall- and temperature-driven abundance model for Aedes albopictus populations* », la

SAGEO'2018 – Montpellier, 6-9 novembre 2018

*dynamique de population du moustique « tigre » Aedes Albopictus a pu être modélisée pour la première fois en climat tempéré au travers d'une approche mécaniste. La dynamique du modèle est pilotée par les variables météorologiques de température et de précipitation. Cette approche a été validée par comparaison avec des données entomologiques relevées sur quatre années dans la région de Nice (coefficient de pearson 0,73-0,93). La validation satisfaisante du modèle en climat tempéré s'explique principalement par la prise en compte de la diapause, une période défavorable au développement du moustique, et pendant laquelle seuls les œufs des moustiques survivent. Dans un second temps, ce modèle a été spatialisé et implémenté à La Réunion en 2015 dans le cadre du projet ALBORUN, avec pour objectif le développement d'un outil opérationnel à destination du service de lutte anti-vectorielle (LAV) de l'Agence de Santé Océan Indien (ARS OI). Les prédictions du modèle ont montré un bon accord avec les observations de terrain, ce qui a conduit à la construction d'un outil opérationnel de cartographie des densités de moustique intégrant les données de stations météorologiques in situ. Cet outil est actuellement utilisé en routine par les services de LAV à La Réunion. La généralité de ce modèle mécaniste permet aussi d'envisager son application sur l'ensemble des aires du territoire métropolitain où l'espèce Aedes albopictus est installée. C'est l'objet du projet ARBOCARTO, financé par la Direction Générale de la Santé (DGS) et les Agences Régionales de Santé (ARS Auvergne-Rhône-Alpes, Occitanie et Nouvelle-Aquitaine). Dans cette nouvelle étape, l'outil est à nouveau paramétré pour servir de démonstrateur sur les trois sites pilotes de Montpellier, Grenoble, et Bordeaux. L'intérêt des sorties cartographiques des densités vectorielles d'Aedes albopictus est en cours d'évaluation, pour répondre à terme à une double intégration : (i) comme support décisionnel par les ARS mentionnées, et ii) comme support sur le terrain par les Ententes Interdépartementales de Démoustication (EID).*

*Les possibilités d'adaptation du modèle aux différentes aires de distribution reposent en grande partie sur l'étude de sensibilité réalisée dans l'étude publiée. Celle-ci a permis d'identifier en particulier un paramètre clef dans sa contextualisation géographique : la variable dite de « capacité de charge de l'environnement ». Cette variable reflète la capacité de l'environnement à offrir des conditions favorables à la présence de gîtes de ponte pour les moustiques. Dans les projets ALBORUN et ARBOCARTO, le paramétrage du modèle a pu être affiné en établissant une valeur de capacité de charge à partir de données issues de la littérature et des échanges réalisés avec les acteurs de terrain de la démoustication. L'estimation de cette valeur de « capacité de charge de l'environnement » bénéficie de l'intégration de données géographiques complémentaires pour caractériser l'environnement en termes d'occupation (p.ex. densité de bâti) et d'usage du sol (p.ex. hotspot type cimetière). Une meilleure estimation de la capacité de charge est également attendue avec l'intégration d'informations issues des données de télédétection, en particulier avec l'estimation de la densité de végétation dans la maille élémentaire du modèle. Des développements méthodologiques portant sur une meilleure intégration des données de télédétection, typologie urbaine en particulier, sont également en cours dans le cadre du projet APUREZA (« Analyses des relations entre paysages urbains dengue et Zika », financement TOSCA CNES 2017-2020). Enfin, une nouvelle perspective d'application est envisagée en Asie du Sud-Est avec le projet ECOMORE II (ECONomic development, ECOsystem MODifications, and emerging infectious diseases Risk Evaluation), financé par l'Agence Française de Développement.*

*MOTS-CLES : Aedes albopictus, arbovirus, densité vectorielle, modélisation, température, pluviométrie, télédétection, acteurs*



---

# Vers la conception d'une action de réponse pour la gestion de catastrophe, utilisant la modélisation de connaissances

Claire Prudhomme<sup>1</sup>, Ana Roxin<sup>2</sup>, Christophe Cruz<sup>3</sup>,  
Frank Boochs<sup>4</sup>

1. *i3mainz University of Applied Sciences, Lucy-Hillebrand-Str. 2, 55128 Mainz, Germany*

*claire.prudhomme@hs-mainz.de*

2. *Le2i FRE2005, CNRS, Arts et Métiers, Univ. Bourgogne Franche-Comté, Bâtiment i3M rue Sully, 21000 Dijon, France*

*ana-maria.roxin@u-bourgogne.fr*

3. *Le2i FRE2005, CNRS, Arts et Métiers, Univ. Bourgogne Franche-Comté, Bâtiment i3M rue Sully, 21000 Dijon, France*

*christophe.cruz@ubfc.fr*

4. *i3mainz University of Applied Sciences, Lucy-Hillebrand-Str. 2, 55128 Mainz, Germany*

*frank.boochs@hs-mainz.de*

---

*RÉSUMÉ. Ce travail de positionnement souligne les problèmes actuels liés aux différents aspects de la collaboration inter agence pendant la réponse à une catastrophe. La coordination et la coopération dépendent de l'usage et du partage d'informations, qui doivent faire face aux problèmes d'interopérabilité, de droits d'accès et de qualité. Le projet de recherche a pour but de fournir une évaluation de l'impact de l'information sur la réponse à une catastrophe afin d'aider la prise de décision concernant l'usage et le partage de l'information en répondant à des questions telles que: quelles informations devraient être partagées ou quelle qualité de données utilisées pour améliorer la réponse à une catastrophe? Notre approche de recherche propose de combiner un système d'information capable d'intégrer des données hétérogènes spatiales et un système de simulation multi-agent pour évaluer différentes stratégies de partage, d'utilisation et de dissémination d'information. Une base de connaissance fait la liaison entre le système d'information et le système de simulation. Cette base de connaissance permet la création dynamique de la simulation en fonction de données ouvertes et permet la gestion de la connaissance et des informations propres à chaque agent.*

SAGEO'2018 - Montpellier, 7-9 novembre 2018

*ABSTRACT.* This position paper highlights current problems linked to the aspects of the multi-agency collaboration during disaster response. The coordination and cooperation depend on the information sharing and use which must face up to interoperability, access rights, and quality problems. The research project aims at providing an assessment of information impact on the disaster response in order to support the decision-making about what information shared or what quality of data used to improve the response efficiency. Our research approach proposes to combine an information system able to integrate heterogeneous data and a simulation system to assess different strategies of information sharing, dissemination, and use. A knowledge base is used as a bridge between an information system and a simulation system. This knowledge base permits the design of a dynamic simulation according to open data and for managing the own knowledge and information known by each agent.

*MOTS-CLÉS :* Système de simulation multi-agent, Sémantique, Réponse de gestion de catastrophe, gestion de l'information

*KEYWORDS:* multi-agent based simulation system, semantic, disaster response, information management

---

---

# De la carte au ciel : une approche empirique pour l'étude des stratégies visuelles de pilotes experts

**Raffaella Balzarini<sup>1</sup>, Francis Jambon<sup>2</sup>**

*1. Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, 38000 Grenoble, France  
raffaella.balzarini@inria.fr*

*2. Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG, 38000 Grenoble, France  
francis.jambon@imag.fr*

---

*RESUME. Notre étude s'inscrit dans le domaine de la géo-visualisation, appliquée à l'apprentissage du pilotage des avions. L'étude présentée ici fait partie d'un projet de recherche dans le domaine des Environnements Informatisés pour l'Apprentissage Humain (EIAH). Ce projet a pour objectif le développement de dispositifs d'assistance au formateur, capables d'aider celui-ci dans le suivi de l'activité de l'élève pilote. L'un des verrous de ce projet est le diagnostic de connaissances expertes, et notamment celles mises en jeu lors de la lecture de cartes aéronautiques.*

*L'attention d'un pilote d'avion est constituée de multiples dimensions, mais son principal composant est l'attention visuelle. Celle-ci peut être étudiée grâce aux techniques d'oculométrie qui permettent d'enregistrer les mouvements oculaires d'une personne (Ware 2008). Les techniques et méthodes d'oculométrie appliquées à la cartographie ont vu leur apparition dès les années 70. Des travaux les plus récents se sont focalisés sur des aspects d'usage des cartes interactives, tels que l'efficacité et l'efficacité de la conception cartographique (Coltekin et al. 2010) ou la prise de décision spatiale, mesurés avec des critères d'intensité du regard et/ou de trajectoire oculaire. L'analyse de l'attention visuelle permet également aux chercheurs d'enrichir les modèles cognitifs qui décrivent et prédisent comme l'humain se comporte et raisonne dans l'espace (Keifer et al. 2017). Dans ce contexte, notre étude s'inscrit dans les cadres théoriques et méthodologiques de la recherche en géo-visualisation et facteurs humains (Montello 2002) et s'appuie sur les paradigmes de la recherche en expertise (Ericsson et al. 2006).*

*L'objectif principal de notre étude est d'identifier et modéliser les connaissances « expertes » mobilisées lors d'un vol en condition de « vol à vue » (VFR), c'est-à-dire lorsque le pilote est capable de piloter son avion en gardant une référence visuelle à l'extérieur de l'avion. Plus précisément, notre étude se focalise sur la première tâche essentielle à la conduite d'un vol : le choix des points de repères au sol sur la route prévue. Un repère de navigation (Waypoint) est un repère géographique qui permet notamment au pilote de valider sa direction de navigation. Ces points de repère sont traditionnellement marqués au crayon par le pilote sur une carte aéronautique pendant la préparation de la navigation, afin de pouvoir les reconnaître une fois en vol. Le choix des repères de navigation est une tâche difficile pour un élève-pilote. En effet, lors de la préparation de la navigation, il est courant de choisir de « mauvais » points de repère sur la carte, c'est-à-dire des points de repère qui*

*semblent faciles à identifier sur la carte, mais dont les concrétisations géographiques au sol se révéleront difficilement identifiable une fois en vol. En conséquence, l'absence d'indentification d'un ou plusieurs points de repère au cours d'un vol pourront amener l'élève-pilote à s'égarer. D'un point de vue cognitif, le choix d'un point de repère implique l'établissement d'une correspondance entre l'objet géographique représenté sur la carte aéronautique et l'objet géographique réel dans l'environnement (i.e. : un pont, une rivière...). Afin de mieux cerner le processus de correspondance opéré par les pilotes experts sur les repères de navigation, deux questions de recherche ont été explorées : (1) Quels sont les types d'objets cartographiques choisis (et considérés) par les pilotes comme étant des « bons repères » dans le monde réel ? et (2) Quelles sont les caractéristiques de saillance de ces objets cartographiques ?*

*Notre contribution porte sur la méthodologie exploratoire mise en œuvre pour modéliser le choix des « bons repères », ainsi que sur le protocole expérimental, basé sur des méthodes d'oculométrie combinées à d'autres techniques expérimentales, permettant de valider (en partie) le modèle conceptuel.*

*Afin de répondre à ces questions de recherche, une approche expérimentale, constituée de trois étapes, a été définie :*

*(1) Le recueil a priori des connaissances « expertes ». Il s'agit d'une approche qualitative, basée sur l'analyse d'entretiens de pilotes experts et appuyée par les résultats d'un état de l'art approfondi qui ont permis de dresser une liste des repères a priori pertinents (i.e. villes, lacs, rivières, autoroute, voies ferrées...).*

*(2) La définition d'un modèle théorique. Celui-ci permet de décrire chaque point de repère selon des caractéristiques de saillance contextuelle qui découlent de la combinaison d'une saillance perceptuelle (i.e. forme, couleur, dimensions...), d'une saillance sémantique (i.e. type, rôle...) et d'une saillance structurelle (i.e. règles topologiques et position relative), ainsi que d'une saillance cognitive (i.e. représentations mentales). Ce modèle s'inspire du « Saliency Vector Model » de Caduff et Timps (2008), du « LPNDM » de Fang (2012) et est basé sur les règles de relations spatiales couramment appliquées en géomatique.*

*(3) L'évaluation du modèle. Les principales caractéristiques des repères, décrites par le modèle, ont été évaluées avec un protocole expérimental triangulaire, combinant l'analyse de données visuelles, verbales et de traces d'activités (Holmqvist et al. 2011). Dans un processus cognitif, ceci correspond respectivement (a) à l'acquisition de l'information géographique par l'attention visuelle, analysée avec des méthodes d'oculométrie, (b) à la conceptualisation de l'information traitée avec des techniques d'analyse de données verbales et (c) à l'opérationnalisation de l'information à travers la prise de décision et le marquage sur la carte, analysé à travers la description des traces.*

*Le protocole a comporté l'observation, en conditions réelles de préparation d'une navigation, de 7 pilotes instructeurs (assimilés de fait à des pilotes experts) de l'Aéro-Club du Dauphiné (1 femme et 6 hommes) pendant la préparation de 2 routes imposées sur une carte aéronautique standard. Le protocole a été restreint à une hypothèse expérimentale globale<sup>1</sup> : « les repères sont choisis s'ils ne sont pas ambigus et s'ils se différencient de leurs voisins par les règles topologiques de proximité, d'intersection et d'adjacence ».*

*Les résultats préliminaires de cette étude montrent que les pilotes experts choisissent globalement les mêmes objets géographiques et que ces objets peuvent être décrits selon les règles explicitées dans notre modèle. Ceci paraît particulièrement évident pour les paramètres de la saillance structurelle, où les règles topologiques d'intersection et proximité semblent gouverner le choix des pilotes. L'oculométrie a fourni une compréhension profonde de toutes les zones explorées par les participants. En effet, les pilotes observent beaucoup*

<sup>1</sup> Cette hypothèse globale a été déclinée en quatre sous-hypothèses dans l'article dont est extrait ce résumé (Balzarini et Jambon, 2018)

d'objets géographiques potentiellement aptes à être des repères, mais ils ne les choisissent pas, car ils ne correspondent pas aux « meilleurs » repères de navigation. L'analyse des zones d'intérêt (AOIs) avec la mesure du nombre et de la séquence de visite dans chaque zone, illustre aussi que le premier objet regardé et sélectionné par le pilote sur un tronçon de route est le plus souvent le « bon » repère.

Cette étude a permis d'identifier, essentiellement de manière qualitative, les stratégies de choix de pilotes experts des « bons » repères de navigation. Les approches via l'oculométrie ont été déterminantes pour comprendre finement l'exploration visuelle des experts en fonction de leur prise de décisions. Néanmoins, d'autres approches, incluant notamment des observations en condition de vol réel, ainsi que des analyses comparatives sur les stratégies expertes vs novices, sont à mettre en perspective afin de compléter le design du modèle théorique.

Ce projet a été financé par l'équipex Amiqua4Home (ANR-11-EQPX-0002) et le projet LIG Emergence 2017 AIRBORNE. L'article original dont est extrait ce résumé est « Balzarini R et Jambon F (2018) From Map to Sky: an Empirical Study on Visual Strategies of Expert Pilots, Kiefer, P., Giannopoulos, I., Göbel, F., Raubal, M., and Duchowski, A. T., eds. *ET4S Eye Tracking for Spatial Research*, (doi:10.3929/ethz-b-000222256) p. 64-69, Zurich, Switzerland ». Il est téléchargeable en accès libre à l'adresse : <https://www.research-collection.ethz.ch/handle/20.500.11850/222485>

MOTS-CLES : Représentations cartographiques et mentales, cartes aéronautiques, repères, stratégies visuelles, oculométrie.

---

#### Références bibliographiques

- Caduff D and Timpf S (2005) *The Landmark Spider: Representing Landmark Knowledge for Wayfinding Tasks*. In: T. Barkowsky, C. Freksa, M. Hegarty and R. Lowe (Editors), *Reasoning with Mental and External Diagrams: Computational Modeling and Spatial Assistance*. AAAI Press, Stanford, CA, USA, pp. 30-35
- Çöltekin A, Fabrikant, SI and Lacayo M (2010) *Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings*. *International Journal of Geographical Information Systems*, 24(10), 1559-1575
- Ericsson KA (2006) *Protocol analysis and expert thought: concurrent verbalizations of thinking during experts' performance on representative tasks*. *The Cambridge Hand-book of Expertise and Expert Performance*, pp. 223-241, Cambridge University Press, NY
- Fang Z, Li Q, Zhang X and Shaw S (2012) *A GIS data model for landmark based pedestrian navigation*. *International Journal of Geographical Information Science*, 26(5), 817-838
- Holmqvist K, Nyström M, Andersson R, Dewhurst R, Jarodzka H, Van de Weijer J (2011). *Eye tracking: A comprehensive guide to methods and measures*: OUP Oxford
- Keifer P, Giannopoulos I, Raubal M and Duchowski AT (2017) *Eye Tracking for Spatial Research: Cognition, Computation, Challenges*. *Spatial Cognition & Computation*, 17(1-2)
- Montello DR (2002) *Cognitive map-design research in the twentieth century: theoretical and empirical approaches*. *Cartography and Geographic Information Science*, 29 (3), pp. 283-304

---

# Plateforme logicielle pour l'intégration et la composition de services géospatiaux

Meriem Sabrina Halilali<sup>1</sup>, Eric Gouardères<sup>1</sup>,  
Florent Devin<sup>1,2</sup>, Mauro Gaio<sup>1</sup>

1. LIUPPA, Université de Pau et des Pays de l'Adour, 64013 Pau Cedex, France

*meriem-sabrina.halilali@univ-pau.fr, eric.gouarderes@univ-pau.fr, mauro.gaio@univ-pau.fr*

2. EISTI, 2 boulevard Lucien Favre, 64075 Pau Cedex, France

*fd@eisti.eu*

---

**RÉSUMÉ.** Au cours des dernières années de grands progrès ont été réalisés sur les systèmes d'information géographique et sur le World Wide Web. Les services géospatiaux se sont progressivement développés pour fournir un ensemble de fonctionnalités et de données issues de sources hétérogènes. La question centrale est : comment faciliter, la découverte, l'intégration et l'exploitation de services géospatiaux, susceptibles d'être hétérogènes, complémentaires voire conflictuels, pour en assurer l'interopérabilité et fournir des données métiers pour les géosciences. L'objectif de cet article est la présentation de notre démarche pour la conception et le développement d'une plateforme logicielle qui puisse répondre à cette question. Ce travail s'intègre dans le cadre du projet ANR CHOUCAS pour l'aide à la localisation de victimes en montagne.

**ABSTRACT.** During last years great progress were realized on the geographical information systems and on the World Wide Web. Geospatial services have been gradually developed to provide a set of features and data from heterogeneous sources. The key question is: how to facilitate the discovery, integration and exploitation of geospatial services, likely to be heterogeneous, complementary or even conflicting, to ensure interoperability and provide business data for geosciences. The purpose of this article is the presentation of our approach for the design and the development of a software platform that can answer this question. This work is part of the ANR CHOUCAS project for help in locating victims in the mountains.

**MOTS-CLÉS :** Service géospatial, Taxonomie, Composition de services Web, Web sémantique, Intégration de données géographiques hétérogènes.

**KEYWORDS:** Geospatial Service, Taxonomy, Web services composition, Semantic Web, Integration of heterogeneous geographic data.

SAGEO'2018 - Montpellier, 6-9 novembre 2018

## 1. Introduction

Ce travail s'inscrit dans le cadre du projet ANR CHOUCAS (Olteanu-Raimond *et al.*, 2017). Ce projet implique plusieurs équipes de recherches interdisciplinaires qui contribuent chacune à la création d'une géo-application pour l'amélioration de l'aide à la localisation des victimes en montagne. Un des objectifs est de proposer des méthodes et des outils permettant de développer des processus métiers pour favoriser la prise de décision. Dans notre contexte, un processus métier est une suite de traitements qui vise à constituer et enrichir des données géographiques issues de sources hétérogènes (gazetiers, base de données cartographiques, sites Web de randonnées...) afin de fournir des données métiers aux secouristes et permettre une localisation plus rapide des victimes en montagne.

Les travaux ici présentés proposent la conception et le développement d'une plateforme logicielle mettant à la disposition des partenaires un ensemble de services géospatiaux. Il s'agit de programmes modulaires, ou composants logiciels, conçus pour faciliter la découverte, l'accès et la composition de fonctions de traitement d'informations géospatiales sur Internet. En particulier, cette plateforme intégrera, sous forme de services, différentes sources de données pour en assurer l'interopérabilité et fournir les données métiers. La plateforme est construite autour d'une infrastructure de type SOA (Service-Oriented Architecture). Ce choix a été motivé par les raisons suivantes : favoriser le découpage en services pour la mise en commun des ressources informationnelles ; éviter la multiplication des environnements de déploiement ; éviter l'hétérogénéité des applications et regrouper au sein d'une même application un ensemble de service afin de proposer une vue cohérente ; et enfin disposer du même point d'entrée sans avoir à se soucier de la localisation des centres de traitement et/ou stockage.

À travers cette infrastructure, on s'attache à lever deux verrous scientifiques. Le premier verrou concerne l'intégration de différentes sources de données sous forme de services géospatiaux et la structuration de ces services qui sont susceptibles d'être hétérogènes, complémentaires ou conflictuels. La section 2 présente notre vision pour garantir cette intégration dans un catalogue de services en s'appuyant sur une taxonomie. Le second verrou concerne l'automatisation de la composition des services et la réutilisabilité de ces services composites. L'hétérogénéité et le manque de sémantique dans ces services constituent un obstacle qu'il s'agit de lever. La section 3 explique comment nous envisageons d'utiliser les principes du Web sémantique, notamment la description de taxonomie, pour automatiser le processus de composition de services représentant

des processus métiers. Enfin nous concluons l'article avec quelques perspectives présentées dans la section 4.

## 2. Intégration des services géospatiaux dans un catalogue de services

Il existe une grande quantité de services géospatiaux en ligne qui fournissent diverses données et fonctions géospatiales. La question centrale est de garantir une intégration transparente des sources de données et des fonctions dans notre catalogue, que celles-ci soient internes (des services qui seront développés dans le cadre du projet CHOUCAS), fournies par les partenaires du projet (des services conformes aux standards de l'OGC (Open Geospatial Consortium)<sup>1</sup>, WMS (Web Map Service), WMTS (Web Map Tile Service), WFS (Web Feature Service)) disponibles sur le portail IGN<sup>2</sup>, ou externes en provenance du Web. Afin de faciliter la structuration et l'interrogation de ces derniers, nous utilisons les technologies du Web sémantique. Celles-ci visent à offrir de meilleures méthodes d'intégration, de découverte et d'accès aux données. Elles permettent, en particulier, aux traitements d'exploiter la description sémantique (exprimée selon des standards) des données géospatiales, des services et des chaînes de services. Elles jouent un rôle important également dans la découverte et la structuration des services géospatiaux aux moyens des ontologies et des taxonomies. Nous utilisons ces dernières afin de structurer notre domaine en définissant les concepts importants qui le caractérisent et les relations entre ces concepts.

Dans les travaux qui portent sur les taxonomies pour la description et la classification de services Web géospatiaux nous constatons la présence de deux approches. La première consiste à concevoir des taxonomies génériques basées sur des considérations générales (Bai *et al.*, 2009). C'est le cas de la norme internationale ISO 19119 (19119:2016(fr), s. d.) pour la classification des services Web de l'OGC qui propose une classification s'appuyant sur les principes des architectures logicielles en technologies de l'information (interaction, modèle, traitement, communication...). L'avantage avec ce type de taxonomie, est de procurer un bon référentiel standard et réutilisable. Cependant, elle reste trop générale et souffre de l'absence de description sémantique formelle directement exploitable par une machine. La seconde consiste plutôt à concevoir des taxonomies spécifiques basées sur des considérations métiers (Al-Areqi *et al.*, 2016 ; Yue *et al.*, 2011). Une taxonomie spécifique est en général très bien adaptée au contexte métier auquel elle s'adresse. Sa spécificité la rend malheureusement très peu réutilisable dans d'autres contextes.

---

1. OGC Web services standards, <http://www.opengeospatial.org/standards>

2. portail IGN Institut national de l'information géographique et forestière <http://geoservices.ign.fr>



Pour les raisons citées, et pour aborder le premier verrou nous avons opté pour l'intégration du Web sémantique (Berners-Lee *et al.*, 2001 ; McIlraith *et al.*, 2001 ; Li *et al.*, 2015) pour la description de taxonomies géospatiales. Nous préconisons de produire deux taxonomies distinctes, une pour les services et une pour les données. La taxonomie des services s'appuiera sur une caractérisation fonctionnelle en respectant les normes internationales. Cette taxonomie sera utilisée pour décrire formellement et sémantiquement les services géospatiaux. La taxonomie de données contribuera en particulier à la conception d'un DSL (Domain Specific Language) pour annoter sémantiquement les données d'entrées/sorties des services géospatiaux. Ce DSL sera une spécialisation de langages d'annotation standards tels que TEI<sup>3</sup> et ISO-Space (Pustejovsky, 2017).

Notre approche pour la description sémantique de données et de services géospatiaux vise à améliorer la découverte des services géospatiaux et faciliter la composition automatique de chaînes de services en assurant un bon équilibre entre les taxonomies génériques pour soutenir la réutilisation et spécifiques pour répondre à nos besoins métiers.

### 3. Composition de services géospatiaux pour la création de processus métier

Après l'intégration dans un catalogue de service, nous disposerons de divers services Web géospatiaux (dont le nombre peut être important). Ces services sont susceptibles d'interagir entre eux dans le but de répondre à une requête utilisateur qui correspond à un besoin métier. Considérons par exemple une requête utilisateur qui vise à attribuer une géolocalisation à des entités spatiales dans un texte. S'il existe un service atomique qui permet de satisfaire la requête, celui-ci est directement présenté à l'utilisateur. Dans le cas contraire, il faut construire un service composite. Dans notre exemple, on peut construire une chaîne de services par composition de trois services (Moncla, Gaio, 2017). Le premier pour interroger les gazetiers. Le second pour la désambiguïsation dans la mesure où plusieurs gazetiers sont utilisés. Le troisième pour générer la liste des toponymes.

Pour concevoir de meilleures méthodes de composition de services géospatiaux, nous nous appuyons sur les technologies du Web sémantique. Celles-ci exploitent les descriptions sémantiques de services géospatiaux afin d'en permettre la découverte automatique. La découverte de service consiste à l'émission de requêtes basées sur les concepts présents dans la taxonomie de service, pour rechercher tous les services qui répondent à une certaine fonctionnalité ou besoin métier. Dans un deuxième temps des requêtes sémantiques sont émises selon la description sémantique des données. Elles se basent sur les concepts de la taxonomie données pour assurer la compatibilité des types de données

---

3. TEI norme internationale pour le balisage textuel. <http://www.tei-c.org/Guidelines/P5/>

(entrées/sorties) entre les différents services sélectionnés et assurer leur interopérabilité.

Bien qu'il existe des travaux de ce type (Yue *et al.*, 2007; Al-Areqi *et al.*, 2016), ils ont été développés dans un environnement dédié qui ne favorise pas la réutilisation de services atomiques ou composites. D'autres propositions (Lemmens *et al.*, 2006) en ce sens ont été faites sur la base des architectures SDI (Spatial Data Infrastructure) et des descriptions de services améliorées OWL-S (Web Ontology Language for Services). Bien qu'intéressantes, elles ne prennent pas en considération les standards ouverts actuels pour la gestion des données géospatiales. Ces constats nous incitent à proposer un environnement ouvert qui favorise la réutilisation en s'appuyant sur les standards de l'OGC et les dernières recommandations de l'ISO en termes d'architectures et de technologies logicielles pour les services géospatiaux.

#### 4. Conclusion

Dans cet article, nous avons présenté notre démarche pour l'intégration de services géospatiaux dans un catalogue de services et leur composition pour la création de processus métiers au sein d'une infrastructure de type SOA. Notre démarche se base sur deux taxonomies pour la description sémantique de données et de services géospatiaux. Le but de ces taxonomies est de structurer nos sources de données suivant les caractéristiques communes de ces dernières et de fournir un référentiel commun pour la découverte et la composition des services.

Dans ce contexte, nous avons initié la conception d'une taxonomie spécifique de services géospatiaux à partir des services fournis par nos collègues qui travaillent sur l'appariement de traces GPS avec des descriptions textuelles de randonnées. Sur la base de cette taxonomie, nous avons modélisé la chaîne de services présentée dans (Moncla, Gaio, 2017) en utilisant un éditeur de workflow. L'objectif est de l'implémenter sur plusieurs outils décrits dans la littérature afin d'en évaluer les possibilités. Un premier démonstrateur a été implémenté sur l'environnement jABC<sup>4</sup> qui s'est avéré trop fermé pour nos besoins. Quelques plates-formes d'orchestration de services géospatiaux (ZOO Project<sup>5</sup>, 52° North<sup>6</sup>) ont été suggérées dans les travaux de (Rautenbach *et al.*, 2012). Les plates-formes ont été évaluées pour leurs capacités d'orchestration des services pour la production de cartes thématiques. Nous allons évaluer dans nos prochains travaux ces plates-formes suivant les exigences et les besoins du projet CHOUCAS.

---

4. Java Application Building Center. <http://ls5-www.cs.tu-dortmund.de/projects/jabc/>

5. ZOO Project. <http://www.zoo-project.org>

6. 52° North. [www.52north.org](http://www.52north.org)

Concernant le travail sur les taxonomies, nous allons généraliser notre taxonomie de services pour prendre en compte la classification ISO 19119 et les services fournis par nos partenaires du projet CHOUCAS. La collaboration avec ces derniers doit nous permettre également d'identifier et de caractériser les données géospatiales dont ils ont besoin afin de concevoir notre taxonomie de données.

## Bibliographie

- 19119:2016(fr) I. (s. d.). *Iso/tc 211. (2016). iso 19119:2016(fr), information géographique — services*. <https://www.iso.org/obp/ui/#iso:std:iso:19119:ed-2:v1:fr>.
- Al-Areqi S., Lamprecht A.-L., Margaria T. (2016). Constraints-driven automatic geospatial service composition: Workflows for the analysis of sea-level rise impacts. In *International conference on computational science and its applications*, p. 134–150.
- Bai Y., Di L., Wei Y. (2009). A taxonomy of geospatial services for global service discovery and interoperability. *Computers & Geosciences*, vol. 35, n° 4, p. 783–790.
- Berners-Lee T., Hendler J., Lassila O. (2001). The semantic web. *Scientific american*, vol. 284, n° 5, p. 34–43.
- Lemmens R., Granell C., Wytzisk A., de By R., Gould M., van Oosterom P. (2006). Semantic and syntactic service descriptions at work in geo service chaining. In *Proceedings of the 9th AGILE Conference on Geographic Information Science, shaping the future of geographic information science in europe*, p. 51–61. Visegrad, Hungary, University of West Hungary Press.
- Li H., Wang Y., Cheng P. (2015). Semantic description for the taxonomy of the geospatial services. *Boletim de Ciências Geodésicas*, vol. 21, n° 3, p. 515–531.
- McIlraith S. A., Son T. C., Zeng H. (2001). Semantic web services. *IEEE intelligent systems*, vol. 16, n° 2, p. 46–53.
- Moncla L., Gaio M. (2017). Services web pour l'annotation sémantique d'information spatiale à partir de corpus textuels. In *Spatial Analysis and GEomatics 2017 (Sagéo 2017)*, p. 2–14. Rouen, France. (<hal-01633342>)
- Olteanu-Raimond A.-M., Davoine P.-A., Gaio M., Gouardères E., Van Damme M.-D., Villanova-Oliver M. *et al.* (2017, novembre). Projet CHOUCAS : Intégration de données hétérogènes et raisonnement spatial pour l'aide à la localisation des victimes en montagne. In *Spatial Analysis and GEomatics 2017 (Sagéo 2017)*. Rouen, France. (<hal-01649156>)
- Pustejovsky J. (2017). Iso-space: Annotating static and dynamic spatial information. In *Handbook of linguistic annotation*, p. 989–1024. Van Godewijkstraat 30, 3311 GX Dordrecht, The Netherlands, Springer. (ISBN 978-94-024-0879-9)
- Rautenbach V., Coetzee S., Strzelecki M., Iwaniak A. (2012). Results of an evaluation of the orchestration capabilities of the zoo project and the 52<sup>r</sup> north framework for an intelligent geoportal. *ISPRS Annals of Photogrammetry, Remote Sensing*

*and Spatial Information Sciences*, vol. I-4, p. 163–168. Consulté sur <https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/I-4/163/2012/>

Yue P., Di L., Yang W., Yu G., Zhao P. (2007). Semantics-based automatic composition of geospatial web service chains. *Computers & Geosciences*, vol. 33, n° 5, p. 649–665.

Yue P., Gong J., Di L., He L., Wei Y. (2011). Integrating semantic web technologies and geospatial catalog services for geospatial information discovery and processing in cyberinfrastructure. *GeoInformatica*, vol. 15, n° 2, p. 273–303.

---

# **[Démon] Application mobile Nos pollutions - Le crowdsourcing au service de l'environnement pour caractériser les pollutions visibles**

**Vincent Bonnal<sup>1,2</sup>, Maxime Cousseau<sup>1,2</sup>**

1. CIRAD, UMR TETIS, F-97130 Capesterre-Belle-Eau, France  
Neufchâteau, Sainte-Marie, F-97130 Capesterre-Belle-Eau, France  
vincent.bonnal@cirad.fr, maximecousseau@orange.fr

2. TETIS, Univ Montpellier, AgroParisTech, CIRAD, CNRS, IRSTEA, Montpellier, France

---

*RÉSUMÉ. L'application mobile Nos pollutions a comme objectif de produire, par le grand public, des informations d'observation de pollutions environnementales, notamment de déchets agricoles, mais aussi toutes autres formes de pollutions. Ces observations spatialisées contribueront à l'évaluation des différentes perceptions du grand public vis-à-vis des pollutions présentes en Guadeloupe. Afin de faciliter la création d'un large processus collectif d'observation et de surveillance de l'environnement, le prototype a été développé pour l'environnement Android, afin de permettre aux utilisateurs de transmettre leurs observations d'objets polluants. Les perspectives ouvertes par ce prototype sont nombreuses en termes d'identification et de remédiation de sources de pollution.*

*ABSTRACT. The mobile application prototype Nos pollutions aims to produce, by a large public, informations of observations of environmental pollution, including agricultural waste, but also all other forms of pollution. These spatialized observations will contribute to the assessment of the different perceptions of the general public concerning the pollution present in Guadeloupe. In order to facilitate the creation of a large collective process of observation and monitoring of the environment, the project is to develop a mobile application, under Android, to allow users to transmit their observations of polluting objects. The perspectives opened by this prototype are numerous in terms of identification and remediation of pollution sources.*

*MOTS-CLÉS : pollution, production participative, processus collectif d'observation, surveillance de l'environnement, application mobile, Android, impact environnemental*

*KEYWORDS: pollution, crowdsourcing, collective observation process, environmental monitoring, mobile application, android, environmental impact*

---

SAGEO'2018 – Montpellier, 6-9 novembre 2018

### **1. La pollution diffuse en Guadeloupe fait oublier la pollution visible**

Aux Antilles, le contexte de la pollution par la chlordécone a facilité la prise de conscience de la société civile de l'impact des activités agricoles sur l'environnement et sur la santé humaine. Des projets de recherche se consacrent à la pollution agricole par l'étude des mécanismes de circulation des polluants entre les différents compartiments de ces milieux insulaires, mais également à l'impact des activités anthropiques sur l'environnement. Le projet Rivage intitulé "Innover en agroécologie pour gérer, préserver et restaurer la qualité environnementale du territoire" s'intéresse aux processus physiques (tels que le transfert des polluants entre l'eau et le sol) et aux processus humains afin de mieux connaître les relations complexes entre les pratiques agricoles et leurs impacts environnementaux. Les représentations que se font les différents acteurs des pollutions sont aussi étudiées en faisant une large place à la société civile durant les écoles-acteurs qu'il organise. Ces écoles acteurs sont conçues comme un espace d'échanges, en présentiel, pour susciter des débats avec tous les acteurs de la zone d'étude (Tonneau J-P *et al*, 2017). Durant ces périodes d'échanges, il est apparu que si les perceptions des pollutions diffuses, ou pollutions non visibles, étaient prises en compte, les perceptions du grand public des pollutions agricoles visibles ne l'étaient pas suffisamment. Une des activités du projet a donc été de concevoir un outil permettant d'acquérir massivement des données de suivi environnemental participatif de pollutions visibles d'origines agricoles. Un prototype d'application mobile a été développé et il est présenté dans cet article. L'application a été ouverte aux autres types de pollutions visibles tels que les dépôts de matériaux, d'électroménagers, de déchets verts ou ménagers...

### **2. « Nos pollutions » un outil au service d'un processus collectif d'observation de l'environnement**

Le prototype d'application mobile est nommé "Nos pollutions" et il est à destination du grand public. Il s'appuie sur les principes de la production participative ou crowdsourcing. Ce dernier terme a été inventé par Jeff Howe en 2006 pour définir un processus de production de valeurs, comme peuvent l'être des données ou des informations, reposant sur un large groupe d'individus (une foule) sous forme d'un appel à contribution ouvert voire volontaire. Les technologies mobiles permettent d'accroître le nombre de contributions par rapport aux applications de crowdsourcing classiques sur le Web en les rendant omniprésentes et immédiatement accessibles par les utilisateurs. Elles facilitent la production d'informations grâce aux multiples capteurs embarqués dans les smartphones (Chatzimilioudis G., 2012). Ainsi, l'objectif spécifique de l'application est de rendre possible la création d'un processus collectif d'observation et de surveillance de l'environnement, pour i) faciliter la localisation de toute forme d'objets ou phénomènes polluants visibles, pour ii) comprendre les perceptions du grand public face aux diverses formes de pollutions visibles et pour iii) identifier des possibles

freins d'usage d'un tel outil, qu'ils soient techniques ou sociaux pouvant s'opposer à une diffusion large de l'application finalisée.

Les observations des utilisateurs seront analysées avec des méthodes et outils des sciences sociales et d'extraction de connaissances (tris et analyses statistiques, analyse comparative, fouille de données textuelles,...). Une interprétation du contexte des observations sera réalisée en amont de l'analyse : le contexte local (zone urbaine, rurale,...) grâce à la localisation des observations, et le contexte immédiat (route, champs, bosquet,...) en exploitant les photographies (par photo-interprétation). Les résultats de cette étude seront ensuite intégrés à l'activité du projet Rivage relative aux représentations des pollutions dans l'environnement pour prendre en compte les perceptions du grand public.

### **2.1. Principales caractéristiques**

Le prototype présenté est développé en Java pour un environnement mobile Android. Il s'appuie sur le système d'information (SI) qui constitue l'observatoire territorial de la durabilité des territoires agricoles aux Antilles<sup>1</sup>. Ce dernier a comme objectif de contribuer à l'action publique et de répondre aux enjeux territoriaux (Tonneau J.P. *et al.*, 2017). Pour cela, il organise et facilite le partage des informations et les connaissances à destination des acteurs du territoire sur l'état du milieu, les pratiques agricoles et leur impact sur l'environnement. L'application développée requiert les permissions suivantes : localisation, caméra, accès aux fichiers. Les fonctionnalités principales de l'application sont:

- le stockage et l'affichage des observations déjà réalisées par l'utilisateur sous forme d'une liste et d'une carte dans l'application,
- la synchronisation des observations (entre la base de données distante et la base de données interne) et le téléversement asynchrone<sup>2</sup>,
- la restitution cartographique en temps réel des observations de pollution des utilisateurs sur le SI du projet Rivage.

### **2.2. Collecte des données.**

L'observation d'un objet ou d'un phénomène polluant visible se base sur un ensemble d'informations collectées par l'utilisateur et par différents capteurs du smartphone. Il s'agit notamment de la localisation de l'observation basée sur le relevé des coordonnées du capteur GPS du smartphone, ainsi que sur la photographie de l'objet ou du phénomène polluant observé. L'image prise est redimensionnée afin de ne pas trop consommer de données mobiles et d'espace de stockage. Pour collecter les perceptions de l'utilisateur, l'application propose un questionnaire simple, formalisé par une succession de 3 pages de questions courtes à

---

<sup>1</sup> <https://www.terragriantilles.fr/>

<sup>2</sup> Le transfert des données vers la base de données distante est différé si l'utilisateur a réalisé une observation en zone blanche (zone sans couverture de données mobiles)

choix uniques. Les réponses formulées sont de fait toutes subjectives et dépendent du ressenti et de l'expérience de l'utilisateur.

Ainsi, la première question (Figure 1, image de gauche) concerne la catégorisation de l'objet ou du phénomène polluant observé. La seconde question (Figure 1, image de droite) permet à l'utilisateur d'évaluer l'impact sur l'environnement de ce qu'il observe. La troisième question posée à l'utilisateur est relative à la durée supposée de présence de l'objet ou du phénomène polluant observé. La quatrième question est relative aux possibilités de remédiation. La dernière question concerne le sentiment éprouvé par l'utilisateur qu'il peut exprimer avec ses propres mots (Figure 2, image à gauche). L'observation réalisée est ensuite ajoutée à la liste des observations de l'utilisateur (Figure 2, image à droite) et transmise au serveur.

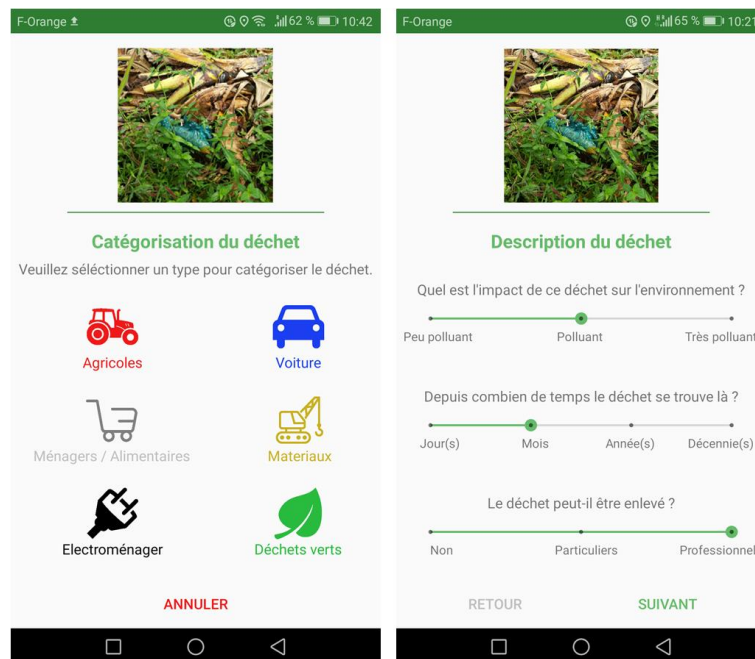


FIGURE 1. Caractériser la pollution observée  
Catégoriser (gauche) - Décrire (droite)

En mettant en perspective le contexte local et le contexte immédiat, les différentes perceptions des utilisateurs, et donc à plus grande échelle celles du grand public, pourront être analysées. Quel est le critère discriminant entre un déchet vert et un déchet agricole lorsqu'il s'agit, pour les deux types d'observations, de matériel végétal ? Quel est le niveau de conscience du public de l'impact environnemental des pollutions dans l'environnement suivant le matériau de l'objet observé (végétal, métal, verre, multicomposants,...) ?



Concernant les sentiments exprimés librement, il pourra être possible d'appliquer des méthodes de fouille de texte sur des sentiments exprimés par les mots et groupes de mots, et de rapprocher les termes utilisés des différentes caractérisations et des contextes.

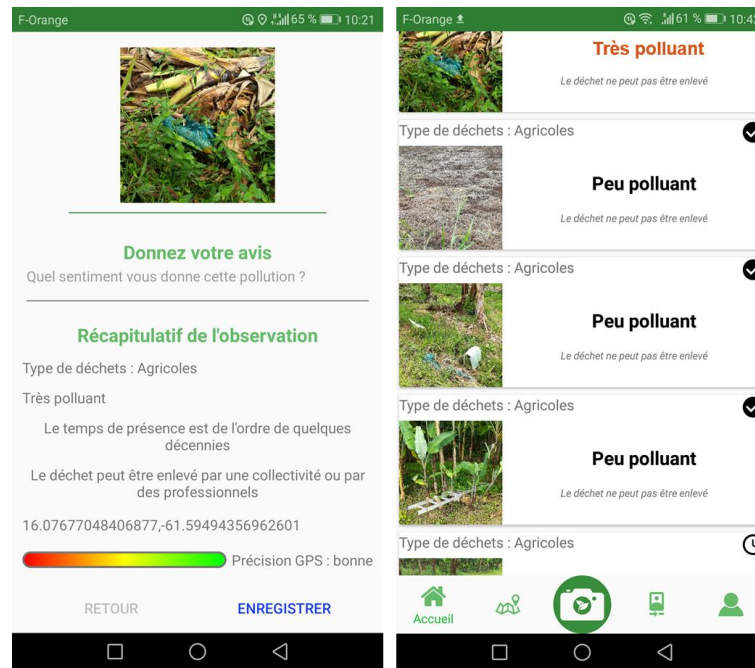


FIGURE 2. *Finaliser l'observation*  
*Donner son sentiment par rapport à une pollution (gauche)*  
*Affichage des observations réalisées (droite)*

### 2.3. Freins sociotechniques et adoption de l'application

La mise en place d'un système de synchronisation différé des données d'observation a permis de supprimer le problème des zones blanches tout en levant un frein à l'adoption de l'application en Guadeloupe en favorisant les connexions wifi, le coût des données mobiles étant bien plus élevés en Guadeloupe que dans l'hexagone. En termes d'ergonomie et d'utilisabilité de l'application, celle-ci a été évaluée par un public de 42 élèves du collège de Petit-Bourg lors d'une sortie sur la zone d'étude du projet Rivage. Les retours ont été très positifs : le principe de l'application a été approuvé par 97% des élèves. 95% d'entre eux recommanderaient l'application à leurs proches. Des propositions de fonctionnalités supplémentaires ont aussi été évaluées. Par exemple, un système de classement des utilisateurs (ranking) a été demandé par 78% des élèves. Il pourrait se reposer sur le nombre d'observations réalisées, mais également sur la fréquence de ces observations.

### 3. Conclusions et perspectives

Le projet Rivage va disposer de nombreuses données afin d'analyser les perceptions du public sur la question des pollutions visibles agricoles. L'application mobile "Nos pollutions" reste à ce jour un prototype, mais l'ensemble du processus de collecte est opérationnalisé. L'application reste encore à perfectionner notamment sur l'optimisation des ressources utilisées (espace mémoire, flux de données,...) avant sa diffusion dans Google Play. Les prochains développements doivent prendre en compte les risques inhérents aux réseaux de production participative par des technologies mobiles (Yang K., 2015) et apporter des garanties techniques quant à la protection des données produites et au respect de la vie privée (Teo SG., 2018).

Pour ouvrir plus largement la collecte des points de vue du grand public, l'ensemble des observations réalisées sera proposé à caractériser et à commenter à d'autres utilisateurs, notamment grâce aux photographies prises. L'objectif de ce partage est de démultiplier les regards sur les objets ou phénomènes de pollutions.

De nouveaux partenariats sont en cours de montage. Les données produites par l'application intéressent le BRGM Antilles pour faciliter l'identification de source de pollutions en surface et qui sont constatées dans les eaux souterraines. L'Ademe souhaiterait pouvoir exploiter les données produites pour accompagner les collectivités locales dans la localisation d'espace à dépolluer (décharges sauvages, carcasses de voiture,...) et dans l'organisation de la remédiation de ces derniers.

*Le développement de ce prototype a été possible grâce au soutien financier du projet Rivage par l'Union Européenne, la Région Guadeloupe et le Cirad.*

### 4. Bibliographie

- Chatzimilioudis G., Konstantinidis A., Laoudias C., Zeinalipour-Yazti D. (2012), Crowdsourcing with Smartphones, *IEEE Internet Computing*, Volume: 16, Issue: 5, Sept.-Oct. 2012, <http://dx.doi.org/10.1109/MIC.2012.70>
- Howe J. (2006). The Rise of Crowdsourcing, *Wired Magazine* (juin 2006), <http://www.wired.com/wired/archive/14.06/crowds.html>
- Teo S. G., Amudha N., J. Cao J. (2018), Privacy-preserving survey by crowdsourcing with smartphones, *4th World Forum on Internet of Things (WF-IoT)*, Singapore, 2018, pp. 647-651. doi: 10.1109/WF-IoT.2018.8355179
- Tonneau J.P., Lemoisson P., Coudel E., Maurel P., Jannoyer M., Bonnal V., Bourgoin J., Cattani P., Chéry J.P., Piraux M., Lestrelin G. (2017). Les observatoires territoriaux. Des outils de la société de la connaissance ? *Revue Internationale de Géomatique*, 27 (3) : p. 335-354. <http://dx.doi.org/10.3166/rig.2017.00035>
- Tonneau J.P., Chéry J.P., Bonnal V. (2017). École-acteurs aux Antilles, un espace de représentation partagée de la question des pollutions agricoles, *Actes du colloque Des outils pour décider ensemble du 26 octobre 2017*, Réseau OPDE, Montpellier.
- Yang K. Zhang K., Ren J., Shen X. (2015). Security and privacy in Mobile crowdsourcing networks: Challenge and opportunities, *IEEE Communications Magazine* (août 2015)

---

## [Démo] Un environnement pour la valorisation de données patrimoniales hétérogènes<sup>1 2</sup>

**André Sales Fonteles, Marie-Noelle Bessagnet, Annig Le Parc–  
Lacayrelle, Christian Sallaberry**

*Univ Pau & Pays Adour / E2S UPPA*

*Laboratoire d'Informatique de l'Université de Pau et des Pays de l'Adour, EA3000  
64000, Pau, France*

*prenom.nom@univ-pau.fr*

---

*RESUME. Nous présentons un travail qui, dans un contexte de promotion du tourisme dans les Pyrénées, vise l'utilisation d'atouts méconnus relatifs au patrimoine culturel matériel (architectural et mobilier) et au patrimoine culturel immatériel. L'objectif est double. Il s'agit tout d'abord de constituer une base de données patrimoniale intégrant des données géoréférencées hétérogènes puis, de proposer des services de consultation et de recommandation de ces informations. Ces services sont dédiés à des applications touristiques Web et mobiles et permettent d'exporter tout ou partie de ces données vers l'Open Data.*

*ABSTRACT. We present a work in a context of tourism promotion in the Pyrenees that aims to leverage the use of unknown assets related to the architectural heritage, furniture heritage and intangible cultural heritage. The project goal is twofold. The first step is to build a database integrating heterogeneous and georeferenced heritage data. Then we aim to offer recommendation and querying services to leverage this database content. These services are dedicated to web and mobile tourism applications and allow exporting all or part of this data to the Open Data.*

*MOTS-CLES : patrimoine culturel, fédération de données, visualisation de données, publication de données, géoréférencement.*

*KEYWORDS: cultural heritage, data federation, data visualization, data publication, georeferencing.*

---

---

<sup>1</sup> Étude réalisée dans le cadre du programme de recherche européen TCV-PYR (2017-2020), financé par l'Union européenne (FEDER) en partenariat avec les régions Occitanie-Méditerranée et Nouvelle Aquitaine

<sup>2</sup> Code source : <https://git.univ-pau.fr/liuppa/TCVPYR/importateur> et Prototype de visualisation des données fédérés : <https://bit.ly/2KSLGG8>

## 1. Introduction

Ce travail a pour objectif la valorisation du patrimoine culturel. Comme évoqué par le rapport (Benhamou et Thesmar, 2011) qui met en exergue les liens étroits entre culture, tourisme et économie, nous positionnons la valorisation du patrimoine culturel au service de l'attractivité des territoires et du tourisme.

Dans le contexte patrimonial, il existe de nombreuses initiatives que nous classons dans le domaine de l'inventaire, d'une part, et de la valorisation, d'autre part. En ce qui concerne l'inventaire, nous pouvons citer les bases de données nationales de référence<sup>3</sup> dédiées au patrimoine culturel matériel (PCM), tels que l'architecture et le mobilier. Nous pouvons également citer l'initiative PCI-Lab<sup>4</sup> dédiée au patrimoine culturel immatériel (PCI), relatant une pratique qu'une communauté reconnaît comme faisant partie de son patrimoine. En ce qui concerne la valorisation du patrimoine, nous avons identifié de nombreux travaux dédiés à des visites de grandes villes ou des applications régionales<sup>5,6,7 et 8</sup>. C'est dans ce contexte que nous participons à la valorisation du patrimoine pyrénéen.

Ainsi, dans le cadre du projet européen FEDER TCVPy<sup>9</sup>, nous travaillons en collaboration avec des chercheurs en Sciences Humaines et Sociales (SHS) sur trois axes. Le premier concerne la collecte de données relatives au PCM, au PCI et aux informations géographiques, pertinentes pour le tourisme dans les Pyrénées. Ces données sont stockées de façon hétérogène et sont difficilement accessibles par le grand public. Le deuxième axe permet l'intégration et la structuration de ces données hétérogènes et géoréférencées dans une base de données commune. Le troisième axe vise la valorisation de ces données via deux applications : l'une de recommandation d'itinéraires composés de points d'intérêts (POI) patrimoniaux et l'autre de diffusion en Open Data.

Dans la section suivante, nous présentons l'architecture de la solution proposée en détaillant la fédération de données patrimoniales hétérogènes et les applications de valorisation de ces données. Nous concluons et détaillerons ensuite les perspectives.

## 2. Architecture générale

L'architecture de la solution proposée (Figure 1) est scindée en deux parties : l'une *Fédération (a)* et l'autre *Valorisation (b)*. La partie (a) est dédiée à

<sup>3</sup> <http://www2.culture.gouv.fr/culture/inventai/patrimoine/>

<sup>4</sup> <https://www.pci-lab.fr/>

<sup>5</sup> <http://patrimoine.alsace/appli-numerique/>

<sup>6</sup> <http://patrimoines.iledefrance.fr/taxonomy/term/3/>

<sup>7</sup> <http://inventaire.aquitaine.fr/actualites/videoguide-limousin-devient-videoguide-nouvelle-aquitaine/>

<sup>8</sup> <http://patrimoines.midipyrenees.fr/cartographie-interactive/randonnees-patrimoine/Aveyron/Millau/grands-sites/>

<sup>9</sup> <http://tcvpyr.iutbayonne.univ-pau.fr/>

l'unification des données patrimoniales hétérogènes issues des applications des partenaires. A cette fin, nous avons conçu une application fédératrice (AF), détaillée ci-après. La partie (b) est dédiée à la valorisation des données unifiées par le biais d'une application touriste (AT), permettant à des touristes d'accéder à des itinéraires selon leur profil, et d'une application (AOD) qui publie les données patrimoniales en Open Data.

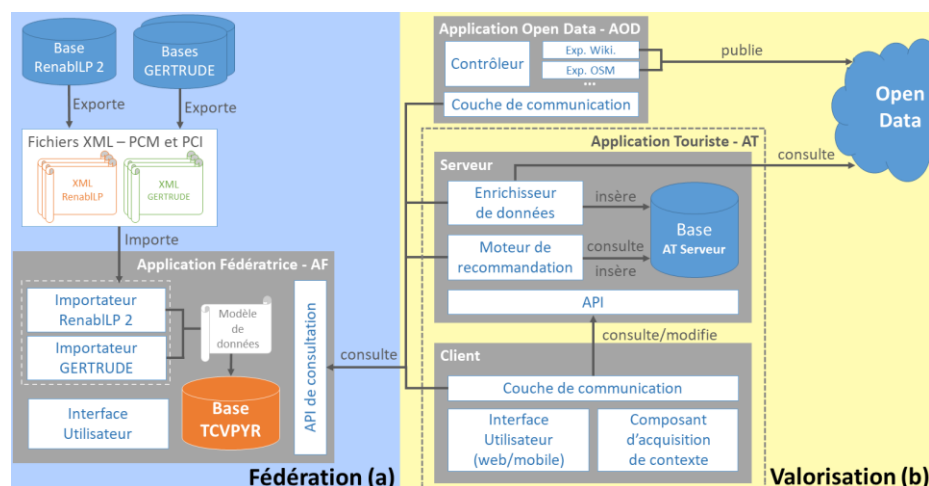
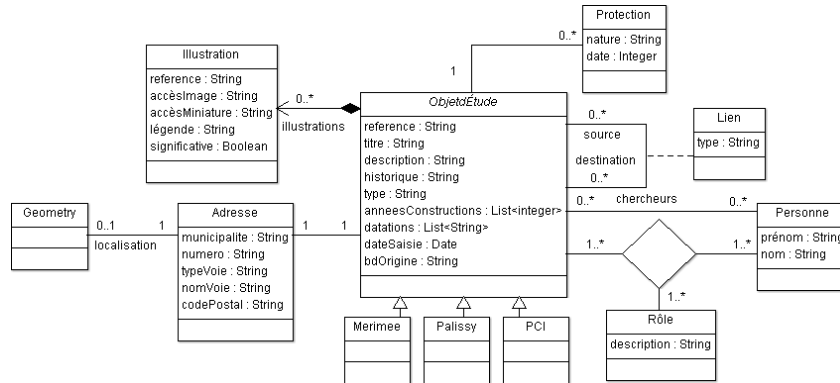


FIGURE 1. Architecture générale

## 2.1. Fédération de données patrimoniales hétérogènes

Afin de collecter les informations patrimoniales, les chercheurs en SHS recueillent sur le terrain les informations qu'ils sauvegarderont par le biais de deux logiciels : RenabLP2 et Gertrude<sup>10</sup>. Ces deux logiciels permettent d'exporter les données collectées au format XML. Cependant, les exports de RenabLP2 et de Gertrude sont très différents aussi bien au niveau de la structuration, de l'organisation que du nommage des informations. Pour exploiter ces données hétérogènes dans un cadre dédié au tourisme, nous avons conçu un modèle de données unifié et homogène (Figure 2). Ce modèle décrit tous les objets d'études des chercheurs : bâtiment, mobilier, PCI. Chaque objet est géolocalisé et possède une description, une histoire ainsi que la liste des personnes qui y sont rattachées (par exemple, l'architecte d'un bâtiment ou le sculpteur d'un objet). Ces objets peuvent être liés : un bâtiment peut être composé de plusieurs bâtiments ou un mobilier peut se trouver dans un bâtiment. Ils sont illustrés par des images. Ce modèle décrit uniquement les informations qui sont utiles pour les applications de valorisation du patrimoine.

<sup>10</sup> RenabLP2 et Gertrude sont des outils de création et de gestion de dossiers électroniques appliqués à la documentation de l'inventaire du patrimoine culturel. RenabLP2 est utilisé en région Occitanie et Gertrude l'est en région Nouvelle Aquitaine et Occitanie.

FIGURE 2 : *Modèle de données*

L'application fédératrice (AF) est incluse dans un processus débutant par des exportations de fichiers XML (issus de RenabLP2 et Gertrude). Ces fichiers sont traités et importés par les composants Importateurs (Figure 1) afin de les rendre conformes à notre modèle unifié. Ces données sont alors stockées dans une base Elasticsearch TCVPYR et exploitées par le biais d'une API de consultation par les applications de valorisation. AF a été développée et est aujourd'hui fonctionnelle.

## 2.2. Valorisation des données patrimoniales

Dans le cadre du processus de valorisation, nous envisageons une application destinée aux touristes (AT) et une autre dédiée à la diffusion des données patrimoniales en Open Data (AOD).

AT exploitera des scénarios de valorisation afin de recommander aux touristes des itinéraires composés de points d'intérêts (POI) patrimoniaux. AT est basée sur une architecture client-serveur (Figure 1). Le client, sous la forme d'une application mobile ou Web, contient un composant de communication avec les serveurs AT et AF. De plus, un composant d'acquisition de contexte de l'utilisateur permettra au serveur la réalisation de recommandations (composant Moteur de recommandation) des itinéraires composés de POI issus de la base TCVPYR et de l'Open Data (composant Enrichisseur). La base AT Serveur stockera des informations nécessaires à la recommandation tel que le profil utilisateur (Picot-Clément, 2011) (Borràs et al, 2014). Nous avons déjà développé un premier prototype de l'AT qui nous permet la visualisation « brute » (sans filtre ou recommandation) des POI de notre base de données (Figure 3). Ce prototype est développé en Java (côté serveur) et JavaScript (côté client). Il interroge la base TCVPYR et affiche les POI obtenus sur une carte via l'API Google Maps.

FIGURE 3. *Prototype de visualisation des POI*

AOD, quant à elle, publiera les données patrimoniales sur l'Open Data, Wikipédia ou OpenStreetMaps par exemple. Nous avons développé une première version du composant pour publication sur Wikipédia. Ce composant est développé en Java. Il interroge la base TCVPYR et publie le POI obtenu sous la forme d'un article par l'intermédiaire de l'API JWiki<sup>11</sup>. La Figure 4 montre un exemple d'article créé automatiquement à partir de l'un de nos POI (le Château Champsaurai).

Le Château Champsaurai est un bâtiment localisé à [Bagnères-de-Luchon](#) et répertorié par la Région Occitanie pour son intérêt culturel, historique ou scientifique.

#### Histoire

Un édifice, qualifié de "castelet" est représenté en cours de construction sur le plan de Luchon par Castex en 1873. S. Liégar (op. cité, p. 77-78) évoque la commanditaire de ce petit château, la séduisante propriétaire de la villa Diana voisine : "près du chalet, sa baguette, mal inspirée cette fois, a fait sortir du sol un castel à tourelles de pierre qui gagnerait à être moins collé au roc : on y travaille encore". Il qualifie par la suite l'édifice de "joujou féodal" pour commenter son style néo-gothique. La commanditaire est Elisabeth Arios, fille du maire de Luchon, Mathieu Soulié, et veuve de Bernardo-Arios. Pour accentuer le caractère médiéval du château, elle se fait créer un emblème (un double E majuscule) qu'on retrouve sur toute la décoration du château (extérieurs, cage et voûte de l'escalier, vitraux, cheminée, mobilier du salon rouge). Le castelet prend par la suite le nom de "château Champsaurai" ("Saurai" est le nom du lieu-dit sur le plan cadastral de 1837) et est photographié sous ce nom par Labouche frères dans la série sur les Pyrénées. Le pyrénéiste Henri Beraldi séjourna autour de 1900 à la villa Champsaurai et fit représenter l'édifice sur une gravure sur bois réalisée par Henri Paillard. En août 1936 le château et son mobilier sont mis volontairement en vente aux enchères (en même temps que la villa Diana) avec une mise à prix à 100 000 F. L'activité hôtelière est relancée l'année suivante où un dénommé de Rolli recrute pour le château Champsaurai du personnel de service.

#### Description physique

Le château est construit en maçonnerie enduite selon un plan proche du carré. Le corps de logis est complété par diverses extensions qui lui confèrent son aspect pittoresque : deux tourelles en encorbellement à l'est, une tour demi-hors-œuvre à pans coupés à l'angle nord-ouest, qui abrite l'escalier et une tour carrée au sud-ouest. Les toitures sont en ardoise. Au nord un escalier extérieur en fer à cheval mène à l'entrée de l'édifice, située en bas de la tour à pans coupés. La porte est surmontée d'un gâble ajouré. L'accès au niveau de soubassement s'effectue par une seconde porte, située sur l'élévation postérieure. Les baies des deux niveaux intermédiaires sont surmontées d'un encadrement à ressaut qui esquisse au centre un dessin d'accolade. Les piédroits sont ornés de colonnettes qui se transforment en moulure dans leur partie supérieure. Ces colonnettes sont dotées de cannelures sur les baies étroites des tours. Les baies du comble sont percées d'une croisée surmontée d'un fronton triangulaire dont le rampant est orné de crochets. Un balcon filant longe le rez-de-chaussée de la façade : il s'élargit dans sa partie centrale. Les pièces de service sont situées dans le soubassement. Le rez-de-chaussée abrite les lieux de réception, les étages les chambres. La porte piétonne reliant le parc à l'avenue de Vénasque est

**Présentation**

Type : Château

Construction : 3e quart 19e siècle

Reconstruction : 3e quart 19e siècle

Commanditaire : Elisabeth Arios,

**Localisation**

Pays : France

100000

FIGURE 4. *Article sur le Château Champsaurai*

## 4. Conclusions et perspectives

Nous avons présenté l'architecture générale de la solution proposée en décrivant la fédération de données patrimoniales hétérogènes et les applications de valorisation de ces données relatives au massif pyrénéen. Le point commun de l'ensemble des données patrimoniales traitées dans le projet TCVPy est leur géoréférencement. Dans les prochains développements, nous envisageons de croiser les profils (position géographique, mode de transport, fenêtre de disponibilité,

<sup>11</sup> <https://github.com/fastily/jwiki>

centres d'intérêts, etc.) des touristes et la base de données patrimoniales pour mettre en place un système de recommandation d'itinéraires de visite comprenant des POI patrimoniaux pertinents. Ces itinéraires seront composés de POI incontournables mais aussi d'un nombre raisonnable de POI insolites. L'application fédératrice et un premier prototype de visualisation de POI ont été rendus publics<sup>12</sup>.

### **Bibliographie**

- Benhamou F., Thesmar D. (2011). *Valoriser le patrimoine culturel de la France*. Conseil d'analyse économique, ISBN : 978-2-11-008595-5, Paris.
- Borràs J., Moreno A., Valls A. (2014). *Intelligent tourism recommender systems: A survey*. Expert Systems with Applications, Volume 41, Issue 16, 2014, Pages 7370-7389.
- Picot-Clément R. (2011). Une architecture générique de Systèmes de recommandation de combinaison d'items : application au domaine du tourisme. Thèse de l'Université de Bourgogne.

---

<sup>12</sup> Code source disponible sur : <https://git.univ-pau.fr/liuppa/TCVPYR/importateur>



---

# [Démon] Regioviz, un outil de visualisation et de comparaison régionale<sup>1</sup>

Matthieu Viry<sup>1</sup>, Ronan Ysebaert<sup>1</sup>, Marianne Guérois<sup>2</sup>

1. UMS RIATE, (CNRS – Université Paris Diderot – CGET)  
8 place Paul Ricoeur, Université Paris Diderot, Bâtiment Olympe de Gouges,  
75013 Paris, France
2. UMS RIATE, UMR Géographie-cités  
8 place Paul Ricoeur, Université Paris Diderot, Bâtiment Olympe de Gouges,  
75013 Paris, France

matthieu.viry@cnrs.fr ; ronan.ysebaert@cnrs.fr ; marianne.guerois@univ-paris-diderot.fr

---

**RÉSUMÉ.** *Regioviz est un outil interactif en ligne qui permet de comparer simplement une région française à l'ensemble des régions européennes. Il répond à une commande de l'Observatoire des territoires du CGET, qui a coordonné la réponse à une demande que les Secrétaires Généraux aux Affaires Régionales (SGAR) ont exprimé auprès du directeur des stratégies territoriales du CGET. Ils souhaitaient disposer d'un outil partagé, simple d'utilisation, leur permettant d'évaluer les effets de la recomposition régionale de 2016 sur un certain nombre d'indicateurs structurels, dans un contexte européen. C'est l'UMS RIATE qui a développé entre 2017 et 2018 cet outil de géovisualisation visant à situer et comparer les nouvelles régions françaises dans un contexte européen : Regioviz.*

**ABSTRACT.** *Regioviz is an interactive tool developed to easily compare French regions to the European ones. It answers to a need expressed by the Observatoire des Territoires (CGET), who has synthesised a request of information of the Secrétaires Généraux aux Affaires Régionales (SGAR) delivered to the director of territorial strategies of the CGET. The SGAR wanted to have on hand a tool, easy to use, allowing to monitor the effects of the regional reform of 2016 on several structural indicators in a European context. UMS RIATE has developed from 2017 and 2018 this geovisualisation tool which aims at positioning and comparing the new French Regions in this context: Regioviz.*

**MOTS-CLÉS :** *Géovisualisation, Outil, Position, Ressemblances, Régions, Europe.*

**KEYWORDS:** *Geovisualisation, Tool, Position, Similarities, Regions, Europe.*

---

<sup>1</sup> Outil disponible à l'URL suivante : <https://riatelab.github.io/regiovizlab>

## 1. Répondre à une demande des acteurs institutionnels régionaux

Regioviz est un outil interactif en ligne qui permet de comparer simplement une région française à l'ensemble des régions européennes. Il répond à une commande de l'Observatoire des territoires du CGET, qui a coordonné la réponse à une demande que les Secrétaires Généraux aux Affaires Régionales (SGAR) ont exprimé auprès du directeur des stratégies territoriales du CGET. Les SGAR souhaitent disposer d'un outil, simple d'utilisation, leur permettant d'évaluer les effets de la recomposition régionale de 2016 sur un certain nombre d'indicateurs structurels, dans un contexte européen. En effet, les dispositifs d'observation développés par les régions leur permettent de bien connaître leur territoire et de répondre ainsi aux conditions de réalisation des différents schémas et diagnostics que leur impose la législation. Parallèlement, la réforme territoriale, mais aussi le mouvement global de mondialisation des échanges, les obligent à s'inscrire et à se positionner dans une compétition internationale. De ce point de vue, les régions ont besoin de repères et de comparaison, non seulement entre régions françaises mais surtout par rapport aux autres régions européennes (Jouen, 2015 ; Antunez et al., 2017).

Pour répondre à cette demande, l'UMS RIATE<sup>2</sup> a développé entre 2017 et 2018 un outil de géovisualisation visant à situer et comparer les nouvelles régions françaises dans un contexte européen : Regioviz (UMS RIATE, 2018).

A partir d'une sélection d'indicateurs mûrement réfléchis (disponibilité des données, objectifs de suivi politique) et de supports graphiques et cartographiques adaptés, Regioviz a été conçu pour répondre à un besoin d'informations nécessaires au suivi des politiques publiques à ces échelles d'observation :

- d'une part, à partir de modules élémentaires d'exploration, il s'agit de caractériser la position relative d'une région en fonction de variables pré-sélectionnées. L'espace de référence pour la comparaison pourra correspondre à l'ensemble des régions européennes ou à un sous-ensemble sélectionné en fonction de différents filtres territoriaux, permettant d'identifier des régions du même type que la région d'étude (en termes de structure de peuplement, de situation géographique...).
- d'autre part, pour ces mêmes indicateurs et espaces de comparaison, l'objectif est de mettre en évidence les régions qui partagent des caractéristiques similaires.

---

<sup>2</sup> Le Réseau interdisciplinaire pour l'aménagement et la cohésion des territoires de l'Europe et de ses voisinages (RIATE) a pour objet le soutien aux recherches portant sur l'aménagement du territoire européen. Sous la cotutelle du CNRS, de l'Université Paris Diderot et du CGET, cette unité mixte de service a notamment pour mission de développer des méthodes et outils pour l'analyse territoriale. Pour en savoir plus sur le RIATE : <http://riate.cnrs.fr>

## 2. Naviguer dans Regioviz

Regioviz est organisé en trois panneaux présentés sur un même écran (Figure 1). La partie de gauche est consacrée aux paramètres d'entrée que l'utilisateur va pouvoir sélectionner (quelle unité territoriale ? Quel indicateur ? Quel espace d'étude ? Quelle maille territoriale ?). Sur l'espace central se trouve le graphique où le territoire sélectionné est mis en évidence au regard des autres territoires (surbrillance jaune). L'espace latéral droit est consacré à la carte. Ces trois panneaux sont interreliés et interactifs : à tout moment de son exploration / analyse, un utilisateur peut décider de modifier les paramètres de sélection et de visualisation, l'objectif étant d'apporter un maximum de souplesse aux parcours potentiels d'utilisation. La Figure 1 restitue les étapes de navigation dans Regioviz et présente les principales fonctionnalités de l'outil. La légende de la numérotation présente dans la Figure est spécifiée ci-dessous :

- [1] Choix d'une unité territoriale. Un territoire de référence doit être choisi pour débiter les analyses de Regioviz (saisie avec autocomplétion ou choix dans un menu déroulant).
- [2] Choix d'un ou de plusieurs indicateurs. Un indicateur doit être sélectionné au minimum. Il est possible de sélectionner jusqu'à 7 indicateurs simultanément.
- [3] Choix d'un espace d'étude. Par défaut, l'ensemble de l'espace d'étude est considéré dans les analyses. Il est possible de restreindre cet espace d'étude à des territoires proches ou présentant des similarités sur des critères statistiques, institutionnels ou géographiques.
- [4] Choix d'une fonctionnalité d'analyse. Les fonctionnalités disponibles (non-grisées) correspondent au nombre de variables disponibles. L'analyse peut alors commencer !
- [6] Les options au-dessus du graphique permettent de visualiser les aides pour chaque mode de représentation (méthode de construction, scénario d'utilisation), de générer automatiquement un rapport contenant des clés de lecture des représentations proposées et de télécharger les données, métadonnées et géométries intégrées à l'application.
- [7] Pour chaque graphique, un menu permet de modifier des paramètres ou des sélections prédéfinies.
- [8] Les options situées au-dessus de la carte permettent d'effectuer des actions sur celle-ci (zoom, sélection simple ou rectangulaire) et rappellent le niveau de disponibilité des données.
- [9] La légende est adaptée à chaque type de graphique et aux couleurs qui en sont issues.
- [10] Le tableau sous la carte rappelle les ordres de grandeur de la distribution des indicateurs sélectionnés (minimum, moyenne, médiane, maximum) ainsi que les valeurs pour l'unité territoriale sélectionnée.

- [11] Les indicateurs, espaces d'étude et maillages territoriaux sont documentés en cliquant sur l'icône.
- [12] Une visite guidée de Regioviz est disponible pour permettre à l'utilisateur de découvrir l'environnement de navigation de l'application.

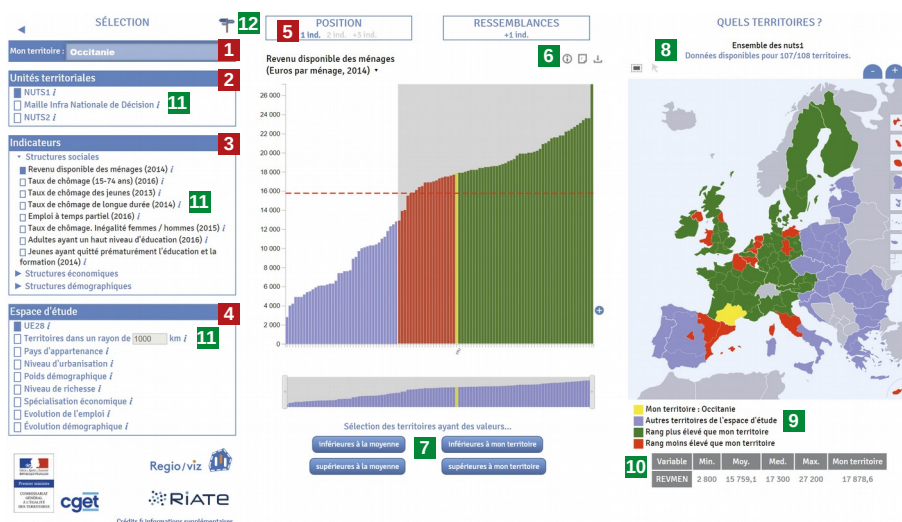


FIGURE 1. Organisation générale de Regioviz

Plusieurs représentations et méthodes sont proposées en fonction du nombre d'indicateurs sélectionnés et du but recherché dans l'analyse (positionner une région, évaluer ses ressemblances avec d'autres). Certaines d'entre elles se retrouvent fréquemment dans les outils de géovisualisation de statistiques régionales (*OCDE Explorer*, *Eurostat cities and regions*)(Jern, 2009), d'autres sont plus originales et ont été créées pour répondre spécifiquement à la demande du CGET de disposer d'un outil simple de comparaison multi-critères, pour au moins 3 indicateurs. (Figure 1 et 2).

Le module « Position – 1 indicateur » propose un graphique en bâtons pour comparer la situation de la région pour un indicateur. Le module « Position – 2 indicateurs » permet d'évaluer la position d'un territoire grâce à un graphique bidimensionnel. Au-delà de trois indicateurs, un graphique en radar permet de comparer la situation du territoire sélectionné au regard de la médiane de l'espace d'étude ou d'autres territoires que l'utilisateur peut sélectionner sur la carte.

Les graphiques de ressemblance permettent de visualiser pour un indicateur et plus les territoires les plus proches statistiquement d'un territoire de référence. Regioviz propose deux niveaux pour la visualisation de ces ressemblances : la ressemblance globale qui permet d'évaluer l'éloignement statistique existant entre

« mon territoire » et les autres territoires de l'espace d'étude au moyen d'un graphique en essaim ; et la ressemblance par indicateur qui permet d'apprécier le degré de similarité entre territoires, indicateur par indicateur.



FIGURE 2. Plusieurs modules d'analyse selon le nombre d'indicateurs et l'angle d'approche privilégié (analyse de position ou de ressemblances)

### 3. Multiplier les angles d'approches et proposer des scénarios d'utilisation

Les principaux destinataires de Regioviz sont des acteurs publics territoriaux qui ont besoin de situer le profil démographique et socio-économique de leur région dans le contexte européen. Si l'un des objectifs est d'aboutir à un outil pédagogique et accessible à un large public, il ne s'agit pas pour autant de mettre ces utilisateurs face à un classement simple et univoque des territoires. Au contraire, une importante marge de manœuvre est laissée pour choisir non seulement les indicateurs mais aussi les découpages territoriaux (différentes définitions possibles de ce qu'est une « région européenne »), les périmètres d'étude plus ou moins vastes (tout ou partie des régions de l'UE28), ainsi que les fonctions d'analyse qui permettent de multiplier les angles d'approche. Pour faciliter la prise en main de ces différentes

fonctions, les utilisateurs sont guidés par une importante documentation qui rend l'interface accessible aux non experts. Celle-ci se déploie à différents niveaux : une description détaillée du projet et des modules d'exploration (aides) ainsi que des exemples de scénarios d'utilisation sont disponibles dans la [documentation de l'outil](#). La fonction de "rapport personnalisé" permet de télécharger les représentations graphiques et cartographiques, accompagnées de textes d'aide à la lecture.

### **Bibliographie non numérotée et références**

- Antunez K., Baccaïni B., Guérois M., Ysebaert R., 2017, « Disparités et discontinuités territoriales dans la France des nouvelles régions : une lecture multiscalaire et multidimensionnelle », in *Economie et Statistique / Economics and Statistics*, n° 497-498, pp. 19-42.
- Jern M., 2009, "Collaborative Web-Enabled GeoAnalytics Applied to OECD Regional Data", *Cooperative Design, Visualization, and Engineering: 6th International Conference*, Luxembourg, September 20-23, 2009, Proceedings, vol. 5738, p. 32-43 [en ligne].
- Jouen M., 2015, « Les nouvelles régions françaises, vues d'Europe », *Policy paper* n°150, Institut Jacques Delors. (<http://www.institutdelors.eu/011-22206-Les-nouvelles-regions-francaises-vues-d-Europe.html>)
- UMS RIATE, (2018). *Regioviz Un outil de géovisualisation pour analyser la position des nouvelles régions françaises en Europe*. Rapport pour le Commissariat Général à l'Égalité des Territoires, 54 p.

---

# Pertinence des SI territoriaux pour la scénarisation énergétique

## *Démonstration par l'exemple de l'exploitation du SI OPTEER appliqué à l'éco-mobilité*

**Julien Crevant<sup>1</sup>, Marie-Hélène de Sède-Marceau<sup>1</sup>, Stéphane François<sup>2</sup>, Massimiliano Capezzali<sup>3</sup>**

1. Laboratoire ThéMA, UMR-6049 CNRS, Université de Bourgogne-Franche-Comté  
32 rue Mégevand, 25000 Besançon, France  
[julien.crevant@univ-fcomte.fr](mailto:julien.crevant@univ-fcomte.fr); [marie-helene.de-sede-marceau@univ-fcomte.fr](mailto:marie-helene.de-sede-marceau@univ-fcomte.fr)
2. ATMO-Bourgogne-Franche-Comté, 37 rue Battant, 25000 Besançon, France  
[stephane.francois@atmo-BFC.org](mailto:stephane.francois@atmo-BFC.org)
3. Pôle Energie et IESE, HEIG-VD  
1 route de Cheseaux, CP 521, 1401 Yverdon-les-Bains  
[massimiliano.capezzali@heig-vd.ch](mailto:massimiliano.capezzali@heig-vd.ch)

---

**RÉSUMÉ.** Dans le cadre de l'ambition REPOS de la région Bourgogne-Franche-Comté, il est primordial de pouvoir scénariser l'évolution de son profil énergétique. Cependant à des échelles infrarégionales, la connaissance et la prise en compte des contextes territoriaux est essentielle pour pouvoir estimer la faisabilité et la pertinence de politiques énergétiques territoriales. Notre étude pose les bases d'une approche multicritère d'aide à la décision en utilisant le système d'information et d'observation OPTEER et ses différentes fonctionnalités ainsi que des données prospectives. L'étude porte sur la mobilité au bioGNV en Franche-Comté et sur l'analyse des impacts qu'aurait une substitution d'une partie du parc de poids lourds.

**ABSTRACT.** Giving the REPOS ambition of the Bourgogne-Franche-Comté, it is necessary to be able to make scenarios in order to plan energy systems evolutions. Yet at regional scale, knowing and taking account of territorial contexts is essential to estimate regional energy policies feasibility and relevance. Our study lays foundation to an aid-to-scenario-making approach using the observation and information system OPTEER, its functionalities and prospective data. The study addresses bioGNV mobility in Franche-Comté and concerns analysis of the impacts a bioGNV substitution in the current truck pool would have.

**MOTS-CLÉS :** OPTEER, aide à la scénarisation énergétique territoriale, territoire, système d'information, éco-mobilité.

**KEYWORDS:** OPTEER, territorial energy scenario making, territory, information system, eco mobility.

---

## 1. Introduction

Ce travail s'inscrit dans le cadre d'une thèse CIFRE effectuée au laboratoire Théma de l'Université de Bourgogne-Franche-Comté (UBFC) en partenariat avec Gaz Réseau Distribution France (GRDF) et la Haute Ecole d'Ingénierie et de Gestion du canton de Vaud (HEIG-VD). Cette thèse a pour but de contribuer à l'unification des ressources informationnelles distribuées et hétérogènes existantes pour poser les bases d'une approche d'analyse et de prospective énergétique territoriale multi-vecteurs et multi-scalaire. Cet exercice de scénarisation nous permet d'aborder plus spécifiquement la problématique de l'aide à la décision multicritère pour la construction et l'évaluation de scénarios régionaux.

La région Bourgogne-Franche-Comté a récemment affiché son ambition de devenir une région à énergie positive (REPOS) à l'horizon 2050<sup>1</sup>. Cette perspective (en plus du cadre légal<sup>2</sup>) impose la mise au point d'une trajectoire cohérente de décarbonation du territoire au travers d'objectifs quantitatifs (qualité de l'air, émissions de GES, consommations énergétiques...). Ce besoin en scénarisation prospective pose un certain nombre de difficultés. Les incertitudes inhérentes à la transition énergétique concernent les évolutions du système énergétique et de son utilisation (décentralisation, technologies, comportements, nouveaux usages, dépendance spatiale et temporelle des productions d'EnR...). Les appropriations locales de la transition<sup>3</sup> mènent à des approches de gouvernance (Magnin, 2013) et d'ingénierie territoriale (Trognon *et al.*, 2012) ou encore au développement de modèles et d'outils spécifiques fonctionnant souvent en boîte noire pour les non-initiés. Les échelles temporelles et spatiales varient également significativement d'une problématique à l'autre (rentabilité économique, temporalités et spatialités des EnR, politiques énergétiques, mandats politiques...) ce qui complexifie l'émergence d'une vision partagée de la question énergétique et de son inscription dans les territoires.

Dans ce contexte, le travail consiste à exploiter les fonctionnalités de la plateforme OPTeER de l'observatoire régional Energie-Climat-Air de Bourgogne-Franche-Comté (ORECA) pour concevoir et diffuser des scénarios régionaux dans le domaine de l'éco-mobilité. L'exploitation de données territoriales via le système d'information (SI) OPTeER répond ainsi à un double-objectif : l'évaluation de la pertinence de trajectoires de décarbonation de la mobilité en regard des opportunités et des contraintes territoriales (appliquée au bioGNV<sup>4</sup>), et la validation de l'intérêt des Systèmes d'Information Territoriaux pour la scénarisation énergétique et l'aide à la décision.

---

1. <https://jeparticipe.bourgognefranchecomte.fr/proposition/faire-de-bourgogne-franche-comte-region-a-energie-positive/>

2. La loi MAPTAM charge la région « d'organiser, en qualité de chef de file, les modalités de l'action commune des collectivités territoriales et de leurs établissements publics pour l'exercice des compétences relatives (...) au climat, à la qualité de l'air et à l'énergie »

3. Conditionnées par les représentations des acteurs du territoire (de Sède-Marceau et Moine, 2012)

4. Gaz Naturel Véhicule d'origine renouvelable



## 2. Périmètre de l'étude

### 2.1. Choix du territoire d'étude

Du fait de son développement en Franche-Comté et de l'historique plus long de collaborations locales, les données de mobilité disponibles dans OPTEER y sont plus riches qu'en Bourgogne. Dans un premier temps, nous ne considérerons donc que le territoire franc-comtois. La Franche-Comté consomme environ 5,7 TWh de gaz naturel par an, ainsi que 12,8 TWh de produits pétroliers pour le transport routier, tous types de véhicules confondus. Si l'on considère les émissions de GES et de polluants plus en détail, on constate sans surprise que la mobilité y occupe une place prépondérante : elle est responsable de près de 50 % des émissions de CO<sub>2</sub>, d'environ 20 % des émissions de PM<sub>10</sub> et de PM<sub>2,5</sub> et de 63 % des émissions de NO<sub>x</sub>.

### 2.2. Choix du bioGNV

Nous concentrons notre étude sur la problématique de la mobilité et sur le vecteur énergétique qu'est le gaz renouvelable sous forme de bioGNV. Le gaz renouvelable utilisable comme bioGNV peut être issu de deux filières : la méthanisation, essentiellement basée sur des ressources agricoles, et la pyro-gazéification, qui utilise du bois (Gassner, 2010). Son principal intérêt est son cycle de production court et local, et donc une empreinte environnementale globale (du puits à la roue) faible par rapport aux produits pétroliers traditionnels (ADEME, 2011). Par ailleurs, au-delà des aspects purement technologiques, la prospective nous impose d'intégrer des éléments liés à la disponibilité et la gestion des ressources, aux concurrences et évolutions des usages, aux spécificités territoriales, à l'évolution des parcs de véhicules, ou bien encore aux contraintes liées aux systèmes de culture et à leur production.

## 3. Mise en œuvre

### 3.1. OPTEER

OPTEER (Observation Prospective Territoriale Énergétique à Échelle Régionale) est initialement un projet de recherche, développé sur des bases transdisciplinaires (voir (de Sède-Marceau *et al.*, 2018) pour une présentation détaillée du dispositif et de sa construction). OPTEER est un dispositif technique d'observation et d'analyse qui s'appuie sur un réseau d'acteurs multidisciplinaires. Depuis 2010, il a contribué à l'élaboration de tous les plans et schémas stratégiques territoriaux portant sur l'énergie, le climat et l'air en Franche-Comté (plus de 80). En plus de l'animation territoriale effectuée par ATMO, la plateforme d'outils d'OPTEER offre un certain nombre de fonctionnalités<sup>5</sup> aux acteurs du territoire, en mettant à leur disposition des

---

5. Qualification et importation de données, hébergement sécurisé, requêtage spatio-temporel, construction de territoires et d'indicateurs personnalisés, visualisation de cartographies thématiques en ligne, création, sauvegarde et diffusion de profils de territoire dynamiques, administration de profils d'utilisateurs, des comptes utilisateurs, des zonages géographiques personnalisés, des données, des formules et des illustrations, mise à disposition d'un portail grand public avec un accès sécurisé aux utilisateurs experts

données énergétiques, des données de qualité de l'air et climatiques ainsi que des données contextuelles.

### **3.2. Données utilisées**

Les données utilisées dans l'étude proviennent de différentes sources et sont pour l'essentiel disponibles dans OPTeER : données de consommations énergétiques ; données climatiques et de qualité de l'air (émissions de GES et de polluants) ; données contextuelles de démographie et de mobilité (source INSEE, ORT-FC...). La scénarisation s'appuie sur des données produites par Solagro<sup>6</sup> dans leur scénario prospectif *Afterres 2050* (Solagro, 2016). Ces données proposent des estimations de gisements potentiellement méthanisable et gazéifiable à horizon 2050<sup>7</sup> (voir annexe pour le détail de ces données).

### **3.3. Méthodologies utilisées**

#### **3.3.1. Construction de jeux de données prospectifs**

Dans un premier temps, nous avons choisi de construire des jeux de données prospectifs simples, afin de pouvoir comparer à l'horizon 2050 les données de gisement potentiel avec les consommations de gaz et de pouvoir estimer ainsi la faisabilité et l'intérêt éventuel du développement du bioGNV.

Les données de gisement méthanisable sont disponibles à la maille cantonale. Celles concernant la pyro-gazéification le sont à une maille départementale. Afin de conserver une lisibilité régionale et de ne pas dénaturer les données, nous nous placerons à une maille départementale. Nous avons choisi d'utiliser les évolutions des consommations de gaz en 2050 issues du scénario Facteur 4 régionalisé de GRDF (hors mobilité). Ce scénario présente des évolutions globales similaires à celui de (ADEME, 2017) et s'appuie sur une solide connaissance du réseau de gaz naturel. La baisse de la demande en gaz prévue par ce scénario (de l'ordre de 35 %) a plusieurs sources, mais nous faisons ici l'hypothèse que les actions de Maîtrise de la Demande en Energie (isolation du bâti, efficacité des technologies...) constituent le moteur principal (et relativement homogène) de cette évolution. Nous supposons également que l'application de ce ratio va permettre de conserver une cohérence à la maille départementale ; cette maille permet de « moyenner » certains contextes territoriaux qui pourraient infirmer cette hypothèse (différences urbain/rural, implications locales plus ou moins marquées...).

Ainsi, nous construisons simplement nos jeux de données prospectifs en appliquant ce ratio aux consommations de gaz actuelles disponibles dans OPTeER. La comparaison de ces consommations prospectives avec le gisement potentiel permet

---

6. Solagro est une entreprise associative reconnue pour son expertise sur les ressources naturelles

7. Ces données sont également utilisées dans l'étude prospective « Un mix de gaz 100% renouvelable en 2050 ? » (ADEME, GRDF, GRTgaz) (<https://www.ademe.fr/mix-gaz-100-renouvelable-2050>)

de calculer les excès ou les manques potentiels pour répondre à la demande en gaz (renouvelable) hors mobilité. Cela donne assez simplement une idée de la marge existante et donc de l'éventuel potentiel de développement.

### 3.3.2. Calcul d'une substitution théorique basée sur le potentiel identifié

Dans un deuxième temps, sur la base du potentiel identifié, nous proposons ici un raisonnement simple permettant de quantifier les impacts qu'aurait la substitution d'une partie du parc actuel de poids lourds fonctionnant au gazole par des véhicules bioGNV. Ce choix s'explique par la répartition en flottes des poids lourds dans les entreprises ou les collectivités, qui permettent des changements de parcs plus faciles à mettre en œuvre que pour les véhicules particuliers, notamment grâce à l'accompagnement de GRDF (connaissances techniques, subventions nationales et régionales...). Les moteurs diesel et GNV étant tous deux thermiques, nous supposons que les avancées technologiques seront comparables, et les calculs sont effectués ici à technologie constante.

Afin d'effectuer les calculs d'impact, nous évaluons la part du trafic interne à la Franche-Comté grâce à la base de données Eider du ministère de la Transition Écologique et Solidaire<sup>8</sup> (MTES). L'application de ce ratio aux consommations actuelles nous permet d'estimer la part qui pourrait être impactée par une substitution. Les données sur le parc de poids lourds sont issues du site Observation et statistiques du MTES<sup>9</sup>. Elles nous permettent d'évaluer le nombre de poids lourds qui serait impacté par la substitution.

Pour les différences d'émissions entre les poids lourds diesel et GNV, nous nous référons au projet Equilibre<sup>10</sup>. Nous y considérons les valeurs de synthèse pour les poids lourds de 44 tonnes, à savoir : une réduction d'environ 10 % des émissions de CO<sub>2</sub>, et une réduction de l'ordre de 60 % pour les NO<sub>x</sub>. La combustion du gaz naturel produit peu de particules fines, et nous ne considérons pas ici les autres sources (usure de la route, des freins et des roues principalement).

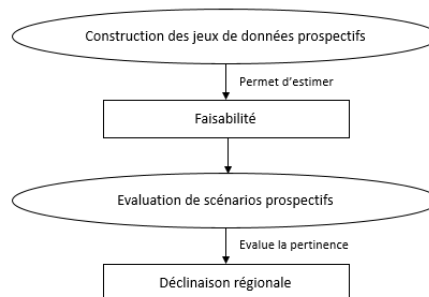


Figure 1 : Démarche méthodologique globale

8. Voir <http://www.stats.environnement.developpement-durable.gouv.fr/Eider/>

9. [http://www.statistiques.developpement-durable.gouv.fr/transports/r/parcs.html?cHash=&tx\\_ttnews%5Btt\\_news%5D=23872](http://www.statistiques.developpement-durable.gouv.fr/transports/r/parcs.html?cHash=&tx_ttnews%5Btt_news%5D=23872)

10. Voir <http://www.projetequilibre.fr/analyse-des-consommations-et-emissions-de-co2-et-nox-sur-des-poids-lourds-gnv-et-diesel/>

Ainsi, nous pouvons calculer la part des émissions actuelles qui pourrait être réduite et estimer simplement la baisse des émissions de CO<sub>2</sub> et de NO<sub>x</sub> correspondant à la substitution diesel/bioGNV. La démarche méthodologique est résumée sur la Figure 1.

#### 4. Résultats

##### 4.1. Faisabilité régionale

L'application de notre méthodologie de construction de jeux de données prospectifs simples nous permet, en exploitant les fonctionnalités de l'outil OPTTEER, de créer une carte (Figure 2) indiquant les potentiels existants en Franche-Comté à l'horizon 2050 (en GWh). On constate ici que la région a un potentiel important, y compris une fois que le gaz dans les réseaux est totalement renouvelable : nous avons ici soustrait la totalité des consommations prospectives de gaz naturel aux potentiels existants, et il reste un potentiel théorique de plus de 7 TWh. Les valeurs extrêmes s'expliquent assez simplement : la Haute-Saône (le plus clair) est un département majoritairement rural et peu dense, tandis que le Territoire de Belfort (le plus foncé) se compose en grande partie de l'agglomération de Belfort et ne contient donc que peu de gisements d'origine agricole ou forestière.

L'augmentation de la part des énergies renouvelables dans les réseaux de distribution est prévue par la loi LTECV, mais celle-ci n'impose pas de verdissement complet d'ici à 2050 (il s'agit cependant du message porté par GRDF, qui s'appuie sur l'étude « 100 % gaz renouvelable en 2050 ? »). On peut donc supposer que le fléchage du biométhane vers une utilisation comme bioGNV permettrait de dégager un potentiel supplémentaire par rapport à une injection dans le réseau. Il faut aussi préciser que cela correspond à un cas où tout le gisement potentiel est utilisé, ce qui paraît peu probable dans la réalité, principalement en raison des usages possibles de la ressource et des arbitrages effectués (par exemple, en Franche-Comté le chauffage au bois est très développé). Ce genre d'analyses « macroscopiques » peut ainsi permettre des discussions autour de l'existence d'un potentiel et d'un choix de développement avant de s'intéresser à sa mobilisation éventuelle et aux bénéfices environnementaux qui en découleraient.

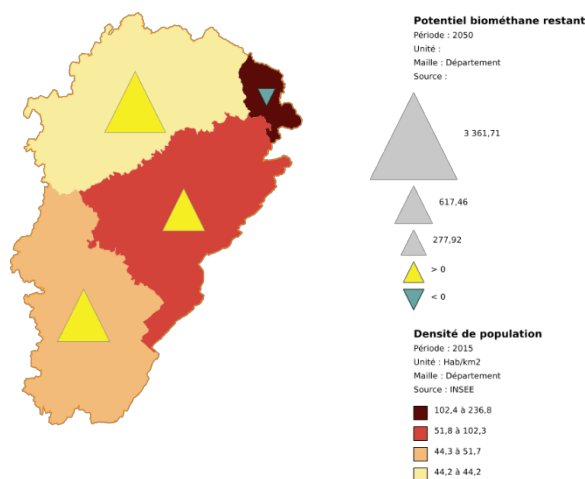


Figure 2 : Estimation du potentiel de production net de biométhane en Franche-Comté

#### 4.2. Substitution de poids lourds et impacts potentiels

La consommation de produits pétroliers des poids lourds en Franche-Comté est de 5,05 TWh, à mettre en regard des 7,2 TWh de bioGNV potentiels. Cependant, cette consommation comprend le trafic interne au territoire ainsi que le transit. Afin d'estimer la part des consommations substituable, il faut donc estimer les parts relatives de ces transports interne et de transit. Les données de la base Eider nous permet d'évaluer la part interne du trafic à 16 %, ce qui correspond à une consommation de 0,808 TWh. Il y a 14 430 poids lourds (bus compris) d'après le site Observation et statistiques, et on évalue à l'aide du même ratio de 16 % le nombre de poids lourds circulant en infrarégional, soit environ 2 300 poids lourds.

OPTEER nous permet de savoir que les émissions dues au transport routier de poids lourds sont de 7 300 tonnes de NO<sub>x</sub> et d'environ 1 277 500 tonnes de CO<sub>2</sub>. L'application du ratio de 16 % à ces émissions nous permet d'estimer la part des émissions qui sera impactée par la substitution vers le bioGNV. On peut ainsi évaluer pour les poids lourds une réduction de 9,6 % pour les émissions de NO<sub>x</sub> et de 1,6 % pour les émissions de CO<sub>2</sub> à l'échelle de la Franche-Comté, soit respectivement 700 et 20 440 tonnes. Les cartes montrant ces gains sont présentées ci-dessous (Figure 4 et Figure 3).

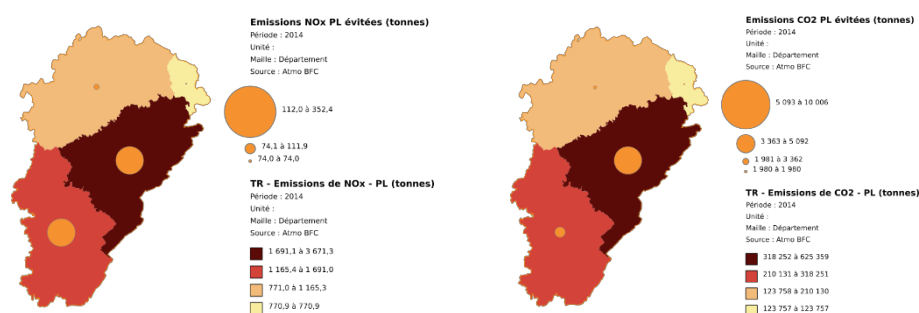


Figure 4 : Emissions de NO<sub>x</sub> évitées par la substitution

Figure 3 : Emissions de CO<sub>2</sub> évitées par la substitution

Ces résultats ne paraissent pas très significatifs en regard des émissions régionales totales du secteur routier. Cependant, dans le scénario proposé, le gaz consommé hors mobilité est entièrement renouvelable. Selon les hypothèses de fléchage retenues, la prise en compte du cycle de vie complet des carburants réduirait donc significativement les émissions de CO<sub>2</sub> dans le cadre d'une substitution. Dans ce cas, le calcul utilisé par Samavati *et al.* (2018) permettrait d'estimer les gains d'émissions sur le cycle de vie complet des carburants en passant du diesel au bioGNV :

$$\text{Gain émissif total} = (E_{\text{fossile}} - E_{\text{bioGNV}}) / E_{\text{fossile}}$$

Cependant, les gains émissifs ne seraient pas tous localisés en Franche-Comté, ce qui rendrait difficile leur spatialisation.

## 5. Discussions et limites

Nous avons vu que l'utilisation d'OPTEER dans un cadre prospectif permet à des utilisateurs non-informaticiens la mobilisation de données externes, la construction de jeux de données prospectifs ainsi que des évaluations simples, visant à débroussailler un futur difficile à prévoir dans le cadre contraignant qu'est celui de la transition énergétique dans les territoires. L'exploitation de données prospectives de gisements a permis l'évaluation macroscopique du potentiel franc-comtois, en nous basant sur des méthodes d'estimation simples, qui demandent à être affinées sur une base plus détaillée, probablement non-linéaire (via la considération d'une typologie de tronçons routiers par exemple, qui impactent les consommations et les émissions). La mobilisation de différentes sources de données a posé un certain nombre de difficultés, notamment quant à leur compatibilité méthodologique et leur finesse : les parcs de véhicules sont au mieux disponible à la maille départementale, ce qui limite les possibilités d'analyse plus fine en l'état.

L'application de la démarche à d'autres territoires permettrait d'estimer la répliquabilité des méthodes utilisées et des résultats, notamment à un niveau national. En outre, il serait intéressant de faire des liens avec les échelles et les mailles plus petites, notamment pour discuter de la déclinaison possible dans un territoire ; la filière méthanisation par exemple pose un certain nombre de questions quant à la localisation des infrastructures et leur raccordement au réseau, quant aux modalités d'approvisionnement en bioGNV pour les camions (stations sur site ? injection dans le réseau et maillage territorial de stations ?) ou encore quant aux données pertinentes dans ce cadre (ressources disponibles, données techniques de réseaux énergétiques, trafic généré par les stations/les installations...).

Les données prospectives et la mise au point d'objectifs peuvent aussi servir de supports de médiation entre les acteurs territoriaux (Noucher, 2009). Les données prospectives pourraient également servir de tiers de confiance dans le cas où (comme ici) les données sont produites par des acteurs extérieurs experts. Cependant les limites posées par les données et leurs échelles de disponibilité (cantonale et départementale dans notre cas) contraignent les possibilités, ce qui souligne l'intérêt d'une meilleure coordination entre les acteurs institutionnels en charge de la prospective et de la décision et les producteurs de données, ainsi que d'une meilleure harmonisation entre les différentes sources de données.

Par ailleurs, les analyses menées à une échelle régionale ne peuvent pas suffire : l'un des intérêts de dispositifs comme OPTEER est de mettre en relation l'information énergétique et territoriale existante avec les acteurs territoriaux (à différentes échelles) connaissant les spécificités et caractéristiques de leur territoire. Dans un cadre prospectif, OPTEER contribue donc à lier les aspects de prospective (scénarisation, long terme) et de planification (mise en œuvre, court terme) via l'implication des acteurs territoriaux.

## Bibliographie

- ADEME (2011). *Analyse du cycle de vie du biogaz issu de cultures énergétiques*, <http://www.ademe.fr/analyse-cycle-vie-acv-biogaz-issu-cultures-energetiques>, Rapport d'études, mars 2011
- ADEME (2017). *Actualisation du scénario énergie-climat ADEME 2035-2050*, <http://www.ademe.fr/actualisation-scenario-energie-climat-ademe-2035-2050>, Rapport d'études, octobre 2017
- Association négaWatt (2017). *Scénario négaWatt 2017-2050*, <https://negawatt.org/Scenario-negaWatt-2017-2050>
- Gassner M. (2010). *Process Design Methodology for Thermochemical Production of Fuels from Biomass. Application to the Production of Synthetic Natural Gas from Lignocellulosic Resources*. Thèse de doctorat, Ecole Polytechnique Fédérale de Lausanne (EPFL), 2010
- Magnin G (2013). La transition énergétique pour quelle société ? International Conference of Territorial Intelligence "Territorial Intelligence, Socio-Ecological Transition and Resilience of the Territories". May 2013, Besançon-Dijon, 2013
- Noucher M. (2009). *La donnée géographique aux frontières des organisations : approche socio-cognitive et systémique de son appropriation*. Thèse de géographie, Ecole Polytechnique Fédérale de Lausanne (EPFL), 2009.
- Trognon L., Cayre P., Lardon S., Maury C. (2012). Ingénierie territoriale : de quoi parle-t-on ? *Productions, Gouvernance et Ingénierie Territoriales*, In Trognon, Baumont, Ingrand, Lardon, Turpin, Vollet (Ed.) Revue d'Auvergne, N°602-603, pp.321-342, 2012.
- Samavati M., Martin A., Santarelli M., Nemanova V (2018). Synthetic Diesel Production as a Form of Renewable Energy Storage. *Energies*. Vol. 11, pp 1-21, May 2013.
- de Sède-Marceau M.-H., Moine A. (2012). Les observatoires territoriaux. Une représentation collective du territoire. *Communication et langages*, n°171, pp 55-65, 2012.
- de Sède-Marceau M.-H., François S., Pauc B. (2018). OPTTEER, une expérience de construction collective d'un dispositif de connaissance et d'analyse territoriale par et pour les acteurs de la transition énergétique. *Numéro spécial de la Revue Internationale de Géomatique « Observer les usages collectifs de l'information géographique »*, vol. 28, n°1, janvier-mars 2018
- Solagro (2016). *Afterres 2050, le scénario 2016*, [http://afterres2050.solagro.org/wp-content/uploads/2015/11/Solagro\\_afterres2050-v2-web.pdf](http://afterres2050.solagro.org/wp-content/uploads/2015/11/Solagro_afterres2050-v2-web.pdf)

## Annexe : Source des données de gisement utilisées et hypothèses structurantes

Il s'agit des données de **potentiel de gaz renouvelable injectable** dans les réseaux en **2050**, donc **après application d'un rendement** de conversion par filière, en TWh PCS

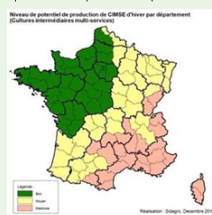
### Rendements considérés:

- méthanisation: **94%** (déterminé par méthane injectable (PCS)/biogaz produit (PCS))
- pyrogaséification: **70%** (déterminé par méthane injectable (PCS) / intrants (PCI))
- power-to-gas: **66%** (déterminé par méthane injectable (PCS) / électricité consommée)

### Intrants agricoles

- Source: *Etude « Afterres 2050 » de SOLAGRO, 2016*
- Maille cantonale
- Pas de concurrence avec les usages alimentaires
- Suppose de *nouvelles pratiques agrologiques*, en particulier la généralisation des *CIMSE*
  - *Pour les CIMSE, il est très important de noter que seul 50% des terrains agricoles où il est possible de mettre en place ces cultures intermédiaires, et qui dépassent un niveau minimum de rendement à l'hectare, ont été considérés dans l'étude.*
  - *Généralisation des CIMSE mais seulement dans les régions propices. Ainsi certaines régions en sont totalement dépourvues du fait d'un contexte pédoclimatique défavorable*

Valeurs à 2050	Masse produite [MtMS]	Masse utilisée [MtMS]	Taux de mobilisation	Energie produite, [TWh <sub>PCS</sub> ]	Milliards de m <sup>3</sup> de méthane
Déjections d'élevages	15	10	67 %	27	2,5
<b>Cultures intermédiaires</b>	41	21	<b>51 %</b>	51	4,6
Herbe et fourrages	53	5,8	11 %	13	1,2
Résidus de culture	62	13	21 %	31	2,8
<b>TOTAL</b>	-	<b>50</b>	-	<b>123</b>	<b>11,2</b>



### Biodéchets

- Sources:
  - *Etude, ADEME, SOLAGRO et INDDIGO, « Estimation des gisements potentiels de substrats utilisables en méthanisation », 2013*
  - *Etude ADEME, GRDF, GREENBIRDIE, « Evaluation du potentiel de production du biométhane à partir des boues d'épuration des eaux usées urbaines », 2014*
- Biodéchets des ménages, de la restauration, des commerces et des industries
- Seulement 46% du gisement net disponible est considéré dans l'étude (7,8 TWh vs 17,4 TWh)

### Biomasse bois énergie

- Sources:
  - *ADEME, IGN, FCBA, « Disponibilités forestières pour l'énergie et les matériaux à l'horizon 2035 », 2016*
  - *Etude INRA et IGN, « Quel rôle pour les forêts et la filière forêt-bois française dans l'atténuation du changement climatique ? » Juin 2017.*
- *Extrapolation au-delà de 2035* en se basant sur le scénario « sylviculture dynamique » établi par ECOFOR
- Sur toute la quantité de bois disponible (usages matériaux & énergie), seul 75% est considéré comme technico-économiquement disponible (accès aux parcelles, pentes, coûts de mobilisation...)
- Pas de concurrence avec les usages « bois matériaux »
  - Uniquement les parties « hors troncs » et les résidus des usages matériaux
- Suppose de la mise en place d'une *sylviculture dynamique*
- *Pas d'arbitrage* de la biomasse entre l'usage chaleur et l'usage gaz
  - Potentiel total disponible pour la production de chaleur et de gaz



---

## Vers des cartes de flux spatialement explicites : la route du niébé ouest africain

Gabriel Poujol<sup>1</sup>, Thierry Joliveau<sup>2</sup>, Jacques Imbernon<sup>1</sup>, Frédéric Lançon<sup>3</sup>

1. UMR TETIS, Cirad

500 rue Jean-François Breton, F-34 093 Montpellier Cedex 5, France

[gabriel.poujol@gmail.com](mailto:gabriel.poujol@gmail.com)

2. UMR EVS, Université Jean Monnet Saint-Etienne/Université de Lyon

6 rue Basse-des-Rives, F-42 023 Saint-Etienne Cedex 02, France

3. UMR ARTDEV, Cirad

73 rue Jean-François Breton, F-34 093 Montpellier Cedex 5, France

---

**RÉSUMÉ.** Cartographier des flux est un compromis délicat entre exhaustivité et pédagogie qui soulève un verrou : comment représenter de façon synthétique et réaliste des mobilités multiples ? Parmi les productions vivrières ouest-africaines, le niébé est une légumineuse principalement produite au Sahel et fortement consommée en milieu urbain. Il est très mobile mais sa circulation est méconnue. Comprendre ce type de mobilités est pourtant essentiel pour aider à la décision dans de nombreux domaines : aménagement, transport, agriculture, intégration économique, sécurité alimentaire. Nos cartes caractérisent les échanges de niébé entre Burkina Faso et Ghana, les excédents et déficits locaux, ainsi que les dépendances spatiales et les surcharges du réseau routier. Nous présentons une méthode de géo-traitement innovante qui génère des cartes de flux spatialement explicites. À partir d'une matrice origines/destinations, nous calculons les chemins des transactions sur les routes et agrégeons ces échanges en densité de flux sur les routes parcourues.

**MOTS-CLÉS :** flux, cartographie, méthode, explicite spatialement, routes, Openstreetmap, niébé, itinéraire, agrégation, Dijkstra, matrice O/D, circuits vivriers, Afrique de l'Ouest

**ABSTRACT.** Flow mapping is a tricky compromise between comprehensiveness and pedagogy that raises a question: how to build synthetic and realistic representation of multiple mobilities? Among West-African staple food crops, cowpea is a legume mainly produced in the Sahel heavily consumed in urban areas. It is versatile but its circulation is unappreciated. However, assessing these mobilities is essential to help decision-making in many areas: planning, transport, agriculture, economic integration, food security. Our maps characterize cowpea trade between Burkina Faso and Ghana, local surpluses and deficits as well as spatial dependencies or road network overloads. We present an innovative geo-processing method that builds spatially explicit flow maps. From an origin/destination matrix, we route transactions on road network and aggregate exchanges in flows density on the roads traveled.

**KEYWORDS:** flow, mapping, method, spatially explicit, roads, Openstreetmap, cowpea, itinerary, aggregation, Dijkstra, O/D matrix, staple food circuits, West Africa

SAGEO'2018 – Montpellier, 6-9 novembre 2018

## 1. Introduction

La cartographie des flux est un exercice délicat, principalement conditionné par la nature des mobilités que l'on souhaite représenter dans l'espace. Les mobilités des biens ou des personnes sont propices à une spatialisation « explicite » des flux alors que des transferts de capitaux ou d'informations par exemple font appel à des représentations de flux dématérialisées. De nombreuses méthodes existent pour représenter des échanges (Bahoken, 2016) et aucune n'est universelle ni parfaite : chacune répond à un besoin spécifique, scientifique ou opérationnel par exemple. Le problème est simple : la représentation attendue doit être synthétique et pédagogique et rendre compte de l'ensemble des processus de mobilité spatiale que l'on souhaite comprendre. Dès lors, représenter des flux sur une carte devient un compromis d'ordre cognitif, selon que l'on vise la synthèse ou l'exhaustivité, la quantité ou la qualité des détails choisis. L'analyse des flux dans l'espace conduit souvent à faire primer la représentation des volumes sur le respect des chemins spécifiquement parcourus. Ainsi, des cartes de flux présentent parfois des échanges qui cheminent hors-sol, empruntant des trajectoires virtuelles, décrochées de la réalité infrastructurelle par souci de lisibilité. Par exemple, de nombreuses représentations cartographiques des flux de commerce bilatéraux ou des migrations pendulaires sont réalisées en distances euclidiennes ou en s'appuyant des chemins, certes simplificateurs, mais irréalistes comme les détroits ou les pôles. Ces représentations hors-sol ne font pas la majorité mais elles révèlent un choix cartographique légitime.

Dans cette communication, nous proposons une méthode de cartographie adaptée, aux mobilités des produits vivriers ouest africains, entre autres, dans un contexte de rareté et de précarité des données. Leurs mobilités restent mal mesurées (Soulé et Gansari, 2010) et très peu spatialisées (Poujol, 2017) alors que la spatialisation explicite des circuits vivriers est primordiale pour renforcer la sécurité alimentaire, l'intégration économique régionale et développer les territoires (Poujol, 2016) et qu'elle permet de surcroît d'appuyer les politiques publiques de transport, d'aménagement et d'agriculture par exemple.

L'exemple choisi porte sur la circulation du niébé, une culture vivrière marchande ouest-africaine, dans le corridor de transport Ouagadougou - Accra, qui relie le Burkina Faso au Ghana. Entre ces deux pays, le niébé transite entre la zone sahélienne burkinabè, où cette légumineuse est majoritairement cultivée, et les villes ghanéennes, notamment en zone littorale. Il s'agit donc de mobilités quantifiables et matérialisables sur un réseau de transport, comme le sont des flux de personnes ou de marchandises en général.

La méthodologie développée permet une représentation cartographique spatialement explicite des flux sur la base d'une matrice Origines/Destinations désagrégée (Bahboub et Morency, 2013) représentant les échanges. À partir d'une matrice O/D issue d'enquêtes sur le commerce du niébé (Poujol, 2017), nous calculons les itinéraires routiers en nous appuyant sur Openstreetmap. Nous encapsulons ensuite les itinéraires obtenus au sein du réseau de routes. Pour obtenir

des flux agrégés, quantifiables, et orientés, nous utilisons une chaîne de géotraitements qui préserve les attributs géométriques et thématiques des itinéraires discrets du niébé.

## 2. Données

Pour cartographier des flux, les données qui décrivent les échanges doivent être formalisées. À partir d'une matrice O/D carrée (*Table 1, gauche*), la transposition en lignes permet d'obtenir la table attributaire d'une couche de points (*Table 1, droite*) représentant tous les lieux (origine, transit éventuel, destination), et les attributs des échanges (nature, quantité, etc.). Les données relatives au commerce de cette légumineuse constituent une couche d'information au sein de laquelle chaque transaction comprend au minimum un couple de points (O/D) dont les attributs thématiques contiennent les mesures discrètes des flux à représenter. Les données relatives aux échanges peuvent être réelles comme simulées, et représenter des volumes ou un simple nombre d'occurrences par exemple.

TABLE 1. *Échanges de niébé, des mesures aux représentations : transposition d'une matrice O/D à la table attributaire d'une couche ponctuelle*

	Vol. niébé	Destinations			FID	ID transac	Pt.	Lib.O	Lib.D	Nat.	Vol.
		Kaya	Wa	Accra							
Origines	Kaya		1000	2500	01	1	O	Kaya	Wa	niébé	1000
	Wa	0		500	02	1	D	Kaya	Wa	niébé	1000
	Accra	0	0		03	2	O	Kaya	Accra	niébé	2500
					04	2	D	Kaya	Accra	niébé	2500

Volumes en tonnes

O : Origines D : Destinations

Les mobilités du niébé présentées Table 1 se traduiront par autant de déplacements le long d'un réseau routier, des points origines jusqu'aux points destinations. Pour les figurer, nous utilisons un module SIG de calcul des plus courts chemins basé sur l'algorithme de Dijkstra (1959). Il reste possible de définir des lieux de transit comme des postes de frontière par exemple pour contraindre l'itinéraire de façon plus plausible ou plus réaliste lorsque c'est pertinent. La base de données routière utilisée pour le calcul des itinéraires est celle d'Openstreetmap, filtrée et corrigée par l'expertise du terrain. Elle est constituée en réseau topologique connecté où le coût de franchissement s'estime en temps de parcours dérivé de la qualité des routes et du mode de déplacement des éléments dont nous analysons la mobilité.

## 3. Méthode

Disposant d'un semis de points Origine-Destination représentant la mobilité des transactions de niébé ainsi que d'un réseau topologique représentant les routes, nous

commençons par calculer l'itinéraire parcouru par chaque cargaison de niébé, et il en résulte autant d'itinéraires que de couples origine/destination, que nous conservons sous forme de polylignes. Chaque polyligne doit disposer d'attributs relatifs à l'échange (origine, destination, nature, volume, etc.) qui décrivent les mobilités que l'on veut agréger, et à sa géométrie (direction azimutale, longueur) qui définissent comment leur superposition permet de les agréger. Ces polylignes sont ensuite fractionnées à toutes les intersections qu'elles ont entre elles, en préservant leurs attributs thématiques et en recalculant leurs attributs géométriques. Cette étape est importante pour préserver la distinction entre deux flux de même nature allant à contresens sur une même portion du réseau par exemple. Elle permet de ne pas les confondre et autorise la représentation de flux bilatéraux. Les tronçons issus du fractionnement des polylignes sont ensuite agrégés en fonction des attributs géométriques qui déterminent leur superposition réelle (même sens) ou virtuelle (contresens) en fonction de la représentation souhaitée, et les mesures relatives aux mobilités qui suivent partiellement les mêmes chemins peuvent être ajoutées.

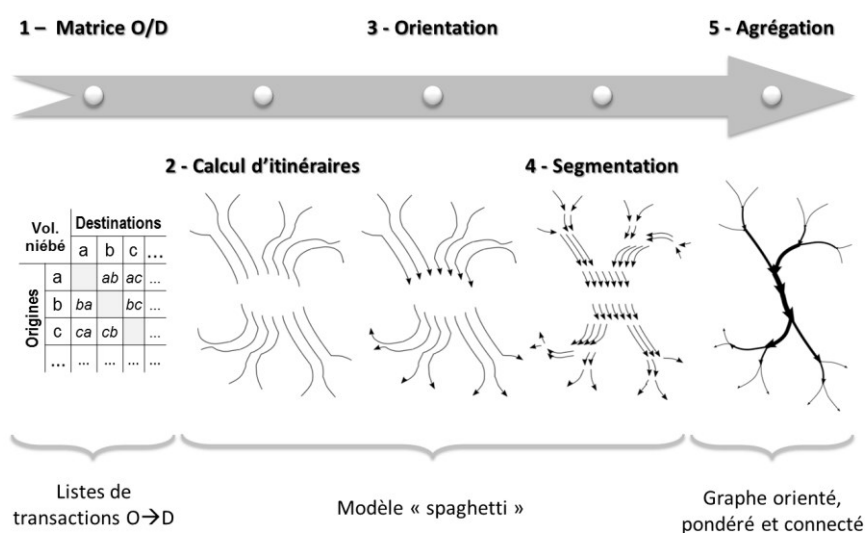


FIGURE 1. Géo-traitements : des matrices d'échanges aux flux agrégés sur les polylignes du réseau routier

Il en résulte un ensemble de polylignes valuées, et possiblement superposées en cas de flux en sens opposés sur un même tronçon. Elles permettent de représenter le flux à travers une expression ponctuelle de sa mesure dans un sens donné (volume de niébé passant par tel point du réseau routier, nombre de transactions, etc.).

#### 4. Résultats

Pour réaliser une cartographie de flux à partir des polygones fragmentées, certaines contraintes doivent toutefois être résolues. D'abord, l'échelle et la représentativité des flux doivent être définies au préalable afin de ne pas garder de volumes trop réduits ou d'itinéraires trop sinueux. Pour cela, nous avons filtré les résultats en éliminant les valeurs agrégées les plus faibles (inférieurs à 2% du flux max.) et en simplifiant la sinuosité des tracés selon un rayon de 5 kilomètres pour limiter la lourdeur visuelle. Ces seuils de filtrage sont arbitraires et doivent être évalués en fonction des contraintes réelles afin de ne pas fusionner ou supprimer des chemins dont la distinction ou la lisibilité seraient capitales. Ainsi, ils améliorent respectivement la significativité et la lisibilité des flux représentés.

À partir des résultats obtenus, il est dès lors possible de cartographier un flux de façon automatique, générique et reproductible<sup>1</sup>, dans les deux sens de chaque tronçon du réseau routier. Il faut appliquer une fonction de décalage du tracé cartographique - existante sous Qgis ou Arcgis - pour que les doubles sens apparaissent, lors de la définition de la symbologie. Bien que plus lourde, une représentation en superposition est également envisageable, si l'on applique des fonctions de transparence par exemple.

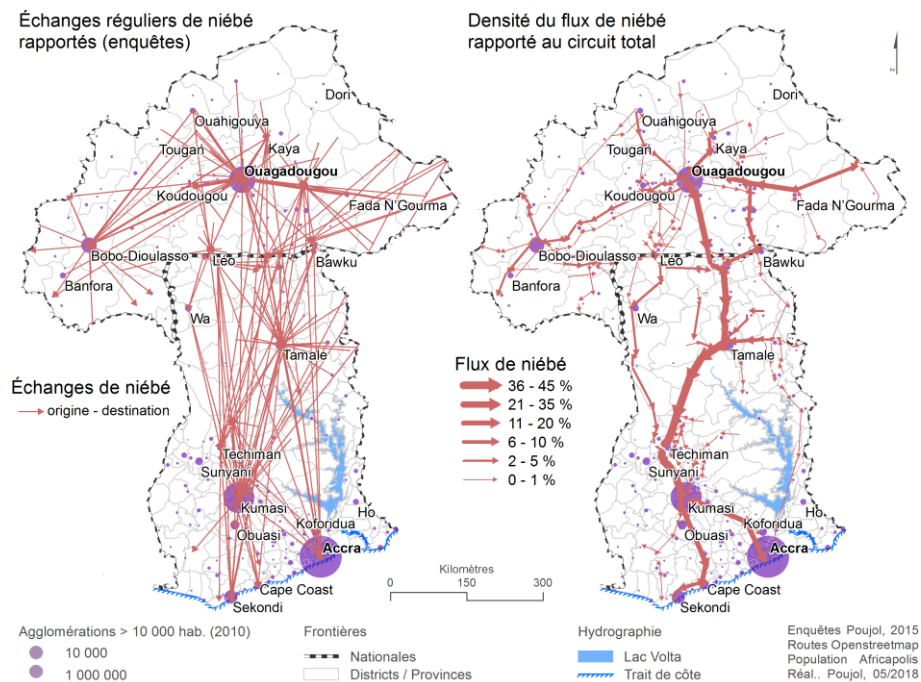


FIGURE 2. La route du niébé du Burkina Faso au Ghana : des échanges aux flux

<sup>1</sup> Modulo un paramétrage des seuils cohérent avec le besoin cartographique attendu.

## 5. Conclusions

Le problème de la cartographie des flux reste donc leur superposition, et la difficulté de les représenter de façon lisible. La conception de modes de représentation n'est pas aisée, tout comme la disponibilité d'objets géométriques pertinents pour figurer leur trace. Par commodité, dans les représentations cartographiques de flux commerciaux par exemple, les mobilités sont souvent décrochées des infrastructures qu'elles empruntent. Les cartes produites représentent ainsi des flux dématérialisés, et deviennent alors peu indicatrices des parcours. Openstreetmap propose une couverture des routes internationale, évolutive et gratuite. Très détaillée dans les pays occidentaux, elle est parfois très perfectible ailleurs mais reste souvent acceptable. Aussi, les outils de géo-traitement disponibles permettent également de réaliser simplement des cartes spatialement explicites qui ouvrent le débat sur un besoin en termes de développement logiciel pour rendre accessible ce type de représentations. Nous pouvons ainsi interpeler la communauté géomatique sur la possibilité d'exploiter l'abondance de données et de géo-traitements à sa disposition pour réaliser des cartes de flux qui proposent un compromis innovant et judicieux entre exhaustivité et synthèse.

## Bibliographie

- Bahboun K. and Morency C. (2013). *Encapsulating and visualizing disaggregated Origin-Destination desire lines to identify demand corridors*, Transportation Research Record n°2430, pp 162-169.
- Bahoken F. (2016). *Contribution à la cartographie d'une matrice de flux*. Thèse en Géographie - Sciences des territoires, Université Paris 7 Diderot.
- Dijkstra E. W. (1959). *A note on two problems in connexion with graphs*, Numerische Mathematik n°1, pp 269-271.
- Poujol G. (2016). *The Ouagadougou Accra Corridor : spatial inclusion through staple food crops*, A New Emerging Rural World n°2, pp 62-63.
- Poujol G. (2017). *Les circuits vivriers du corridor Ouagadougou – Accra : conditions d'un développement inclusif*. Thèse en Géographie et aménagement de l'espace, Université Paul Valéry Montpellier III.
- Soulé B. G. et Gansari S. (2010). *La dynamique des échanges régionaux des céréales en Afrique de l'Ouest*, Michigan State University - Syngenta.

---

# Etude exploratoire des représentations de coulées d'eau boueuse en Alsace : Analyse d'enquêtes psychosociales et de dessins à main levée via un logiciel dédié

**Carine Heitz<sup>1</sup>, Stella Marc-Zwecker<sup>2</sup>, Romaric Mathis<sup>2</sup>, Florence Le Ber<sup>2</sup>**

*1. Irstea, UMR GESTE, ENGEES, 1, Quai Koch, F-67 000, Strasbourg, France  
carine.heitz@engees.unistra.fr*

*2. ICube UMR 7357, Université de Strasbourg, CNRS, ENGEES, F-67 412, Illkirch-Graffenstaden  
florence.leber@engees.unistra.fr; stella@unistra.fr*

---

## RESUME

*Nous étudions la vulnérabilité des territoires face aux coulées d'eau boueuse et plus précisément les représentations par les habitants des paysages exposés à ce risque. Notre hypothèse est que les individus entretiennent une relation non intentionnelle au risque auquel ils sont exposés que nous mesurons via leurs représentations socio-spatiales du paysage. La méthodologie appliquée consiste en une enquête basée sur un dessin à main levée et un questionnaire quantitatif. Le questionnaire permet de collecter des données sociales et le dessin à main levée permet de matérialiser et circonscrire des espaces perçus comme problématiques. Pour écarter la critique sur le caractère trop subjectif de l'analyse des dessins à main levée, nous proposons ici d'utiliser un logiciel que nous avons créé pour en faire un traitement à la fois systématique et informatisé.*

*MOTS-CLÉS : Représentations socio-spatiales, Dessins à main levée, Logiciel d'analyse, Risque érosif, Alsace.*

---

## ABSTRACT

*In this study, we analyze the inhabitants' representations of the landscapes exposed to muddy floods risk. Our hypothesis is that individuals have an unintentional relation to the risk to which they are exposed, which we measure through their socio-spatial representations of the landscape. The methodology consists of a survey based on two different tools: a sketch map and a quantitative questionnaire. The questionnaire allows us collecting social data and the sketch map allows circumscribing territories perceived as problematic. Since sketch maps have often been criticized in the literature for their too subjective analysis, we propose here to use a specific software that we have created to make a systematized and computerized treatment.*

*KEYWORDS: Socio-spatial representations, Sketch maps, Software, Erosion, Alsace*

---

SAGEO'2018 – Montpellier, 6-9 novembre 2018

## 1. Introduction

Les zones des collines lœssiques d'Alsace (Kochersberg, Outre-Forêt, Sundgau, Piémont vosgien) sont très sensibles à l'érosion des sols. La survenue de ces phénomènes érosifs est liée à une conjonction de facteurs défavorables : un climat caractérisé par un maximum d'orages violents en mai-juin ; des sols limoneux à très faible stabilité structurale ; des systèmes agraires faisant la part belle à des cultures peu couvrantes (maïs) au moment des épisodes orageux. La manifestation la plus spectaculaire de l'érosion des sols est la formation de coulées d'eau boueuse qui s'avèrent de plus en plus fréquentes en Alsace depuis ces 30 dernières années (Armand, 2009 ; Heitz, 2009 ; Van Dijk *et al.*, 2005). En effet, une cartographie du risque potentiel de coulée d'eau boueuse en Alsace indique que ce risque est élevé pour 17 % des communes de la région tandis que l'analyse des dossiers communaux de catastrophes naturelles confirme la survenue de ces événements pour 24 années sur 30 dans la période 1985 à 2014 et dans 1 à plus de 20 communes par an (Van Dijk et Koller, 2015). Les coulées d'eau boueuse engendrent des dégradations importantes des agro- et éco-systèmes et sont très coûteuses, notamment par les dégâts occasionnés en zone urbaine, aux infrastructures collectives et à l'habitat individuel.

Les études relatives aux coulées d'eau boueuse s'orientent classiquement autour de l'acquisition de deux types de connaissances : (1) sur l'aléa et les processus physiques en jeu ; (2) sur la vulnérabilité des milieux exposés. Pour diminuer la vulnérabilité, des aménagements (fascines, bassins de rétention, bandes enherbées, *etc.*) peuvent être installés dans le paysage agricole. Ces ouvrages sont la plupart du temps efficaces mais ils modifient le paysage des habitants, souvent attachés à leur cadre de vie.

Notre étude se focalise sur ce dernier point et plus précisément sur les représentations par les habitants des paysages exposés aux risques de coulées d'eau boueuse. Approcher la vulnérabilité par les représentations est une entrée de compréhension du risque par les pratiques et comportements des individus exposés ou sinistrés. Au moyen des représentations, les éléments cognitifs liés à l'individu (perception de son niveau d'exposition, par exemple) et ceux liés à son environnement proche (perception des aménagements, perception paysagère, par exemple) sont identifiés et analysés. Notre hypothèse générale est que les individus entretiennent une relation non intentionnelle au risque auquel ils sont exposés, que nous mesurons via leurs représentations socio-spatiales du paysage (section 2).

Pour répondre à cette hypothèse, nous mobilisons une méthode d'enquêtes particulière, le dessin à main levée accompagné d'un questionnaire. Dans notre cas, le questionnaire permet de collecter des données sociales et le dessin à main levée permet de matérialiser et circonscrire des espaces perçus comme problématiques et de les objectiver sur un dessin (section 3). Etant donné que les dessins à main levée ont souvent été critiqués dans la littérature pour leur analyse trop subjective, nous proposons ici d'utiliser un logiciel pour en faire un traitement à la fois systématisé et informatisé (section 4).



## 2. Les représentations socio-spatiales d'un paysage exposé à un risque naturel : comment les mesurer ?

Pour mesurer les représentations du paysage, nous mobilisons divers cadres théoriques complémentaires : celui des représentations sociales du paysage, celui des représentations du risque et de l'attachement au lieu et celui des représentations socio-spatiales.

### 2.1. Les représentations sociales du paysage comme révélatrices des interactions « individu – environnement »

Les premiers travaux relatifs à l'étude des représentations paysagères des individus datent des années 1970 (Brunet, 1974). Ils s'intéressaient alors aux aménités et aux notions de bien-être rattachées aux environnements familiers des individus. Puis, se sont développées des études sur l'influence de l'esthétique dans les politiques de préservation d'espaces naturels (fin des années 1980 avec Koh, notamment). La majorité de ces travaux se concentre alors sur l'importance d'éléments précis (tels que l'eau, la présence de parcs urbains ou de forêts) dans les représentations de paysages. Ces dernières sont aussi fréquemment mobilisées dans les travaux sur les écosystèmes (Cottet *et al.*, 2010), sur les gestions de milieux aux enjeux environnementaux importants (tels que les bras morts - Le Lay et Piégay, 2007 ; Piégay *et al.*, 2005) ou sur les préférences en termes de renaturation (Caillaud et Marie, 2009 ; Jacopin, 2011). Ainsi, toutes ces études placent les représentations au cœur d'enjeux relatifs à l'utilisation d'espaces « naturels » et à la justification de celle-ci. Elles inscrivent également ces représentations dans un système complexe de relations entre le milieu naturel, les acteurs présents dans ce milieu et leurs interactions avec le paysage (Figure 1).

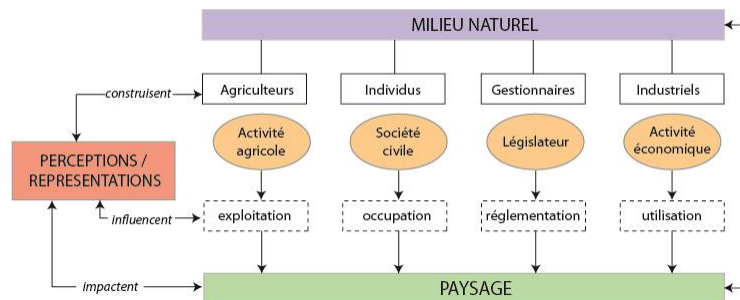


FIGURE 1. Schéma systémique des rapports entre milieu naturel, paysage et perceptions (d'après Jacopin, 2011).

Pour pouvoir qualifier les représentations sociales du paysage, les travaux de Luginbühl (2008) précisent leur articulation autour de 3 échelles traduisant chacune différents niveaux de structuration ou de construction de celui-ci :

- La première échelle détermine des « modèles paysagers » (Cadiou et Luginbühl, 1995) qui sont des « *références symboliques élaborées dans l'histoire des relations sociales à la nature et qui permettent de qualifier un paysage, de le classer dans une catégorie esthétique* ».
- La deuxième échelle est celle de la culture du lieu de vie. Elle intègre les connaissances que se construit l'individu du lieu où il réside. Cette construction se fait par l'observation, le contact ou les autres sens physiologiques de l'humain en relation avec la matérialité de la nature. La mémoire sociale ancre ces connaissances afin que l'individu comprenne le lieu, son histoire, ses mythes et l'origine des formes du paysage.
- La troisième échelle fait référence à l'unicité de l'individu dans l'expérience paysagère et est en lien direct avec sa trajectoire vécue (par ses mobilités, son histoire familiale, par exemple).

Pour chacun des niveaux, l'aspect social marque son empreinte dans la construction de la culture paysagère. D'autres éléments structurants sont également à prendre en compte.

## **2.2. Expliquer les représentations des paysages par les représentations du risque et l'attachement au lieu**

Les travaux de Giddens (1991) marquent le tournant entre la mobilisation de la « société du risque » de Beck (1985) et celle de la « culture du risque ». Aujourd'hui il ne s'agit plus d'éliminer mais de gérer le risque (Peretti-Watel, 2010) en considérant que le « risque zéro » n'existe pas. Il est désormais admis qu'il faut apprendre à vivre avec l'idée d'une exposition au danger (Joffe, 2003 ; Slovic et Peters, 2006 ; Brunet, 2007). Pour étudier le risque, trois approches sont communément mobilisées (Krien, 2014) : (1) l'approche psychométrique du risque qui mesure des processus internes d'analyse dans la même logique individuelle que celle de la perception ; (2) l'approche culturelle du risque (Douglas, 1992) qui étudie les groupes sociaux exclusivement car elle est motivée par une construction sociale « *à partir des valeurs et croyance d'une société* » (Krien, 2014) ; (3) l'approche constructiviste qui unit les deux précédentes car elle prend en compte une construction individuelle du risque influencée par l'environnement social (Kaperson *et al.*, 1988).

L'approche constructiviste nous semble la plus adaptée à notre cas d'étude car « *le risque subjectif consiste en une construction individuelle de la réalité influencée aussi bien par des facteurs sociaux qu'environnementaux* » (Coppiers *et al.*, 2004). Dans cette logique et pour comprendre la relation qu'entretiennent les individus avec leur milieu de vie, de nombreuses études vont alors mobiliser la notion d'attachement au lieu (Bonaiuto *et al.*, 2016). L'attachement se définit comme « *un lien affectif positif qu'un individu ou un groupe développe à l'égard d'un espace donné. Cette notion combine deux composantes : le besoin de contrôle et l'ancrage* » (Krien, 2014). L'ancrage est le besoin d'enracinement, la volonté de s'inscrire dans la durée sur un territoire. Le besoin de contrôle est la familiarisation de l'individu à son cadre de vie pour réduire le stress inhérent à tout changement.

Pour cela, l'individu va marquer le territoire par de multiples aménagements, par exemple (Krien, 2014). Ce sentiment de maîtrise, de contrôle de son environnement permettent alors à l'individu de construire un lien affectif fort avec l'espace dans lequel il évolue (Hidalgo et Hernandez, 2001).

Dans ce même sens, d'autres études indiquent que la forme et l'organisation des représentations sont dépendantes des spécificités socio-culturelles des individus (Jodelet, 1984 ; Ramadier, 2003) d'où la notion de représentations socio-spatiales pour rompre avec cette vision du déterminisme environnemental et s'orienter vers une approche interactionnelle puis transactionnelle de son analyse.

Notre positionnement théorique reprend ainsi trois éléments fondateurs (représentations sociales, attachement au lieu, représentations socio-spatiales) que nous allons mesurer via un protocole d'enquête original et spécifique à la question des représentations des paysages.

### **3. Méthodes d'enquêtes pour collecter les données de représentation des paysages**

La méthodologie appliquée consiste en une enquête basée sur deux outils différents : un dessin à main levée et un questionnaire. Un site d'étude particulièrement exposé aux coulées d'eau boueuse est choisi pour tester ce protocole (la commune de Schwindratzheim, Bas Rhin).

#### **3.1. *Une carte à main levée accompagnée d'un questionnaire quantitatif pour appréhender les représentations socio-spatiales***

Le dessin à main levée est une méthode de collecte de l'information reposant sur la réalisation, sous certaines conditions, d'une représentation spontanée d'un objet ou d'un espace par des individus d'un groupe donné. Le dessin à main levée n'exige pas de se confronter à la réalité, elle peut en être éloignée, avec des formes très variées et de nombreuses possibilités. Cet outil a été développé par Lynch (1960) et repris par d'autres auteurs tels qu'Appleyard (1970), Downs et Stea (1973), Gould et White (1984). L'utilisation du dessin à main levée permet aux individus de reconstituer ou dessiner « leur » paysage quotidien afin d'y distinguer les éléments fédérateurs. En effet, dans ses premiers travaux Lynch (1960) a introduit la notion de « lisibilité » de l'espace en déterminant cinq types d'éléments fondateurs d'un dessin : les quartiers, les nœuds, les points de repères, les limites et les voies.

Les données ainsi obtenues permettent d'identifier les différences existantes entre une représentation et une réalité physique et répondent également à la dialectique entre représentations et usages des paysages. Dans notre protocole, les individus dessinent les éléments paysagers qui leur paraissent importants sur un fond de carte indiquant trois éléments de repères (l'autoroute au nord, un rond-point et au sud un cours d'eau – Figure 2). L'utilisation d'un fond de carte est motivée par les besoins en traitement informatique (cf section 4). Toujours en termes de design expérimental, les individus sont invités à représenter tous les éléments constitutifs de leur cadre de vie en noir et ceux relatifs au risque en rouge. La passation du dessin à

main levée s'est faite sur tablette avec stylet ou papier-crayon afin de mettre en avant des différences entre les deux outils.

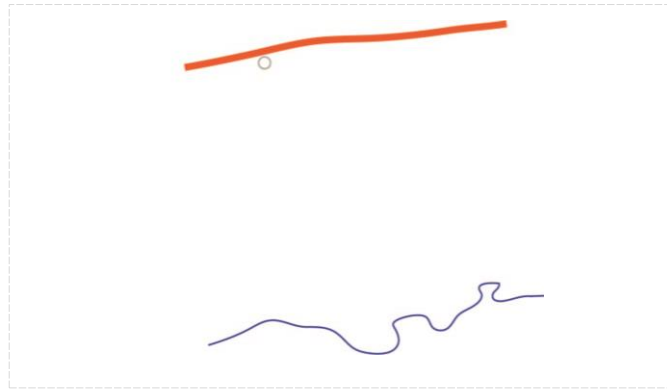


FIGURE 2 : *Fond de carte proposé aux enquêtés lors des passations à Schwindratzheim.*

Le questionnaire permet de collecter des données sociales sur les représentations du risque de coulées d'eau boueuse. Ce questionnaire reprend chaque point théorique développé ci-avant. Nous avons interrogé l'attachement au lieu (10 questions par évocations hiérarchisées) suivant le modèle tripartite de l'attachement au lieu développé par Raymond *et al.* (2010) et celui mobilisé par Krien (2014). Quatre questions (par évocations hiérarchisées) évaluent les connaissances sur le risque de coulées d'eau boueuse (processus et exposition des individus) et les conséquences associées. Trois questions (ouvertes) mesurent les connaissances en termes d'aménagements collectifs pour protéger la commune contre le risque et le sentiment de sécurité induit. Trois questions (échelles de Lickert de 1 à 6) mesurent le sentiment d'information des individus sur le risque. Enfin, dix questions permettent d'obtenir des caractéristiques socio-professionnelles des individus et des informations sur leur trajectoire de vie (notamment en termes de mobilité).

Nous avons divisé la passation en plusieurs temps : le dessin à main levée est proposé en tout début d'entretien pour permettre aux individus de dessiner leur cadre de vie. Puis le questionnaire est proposé. Dans un dernier temps, les individus peuvent compléter le dessin par des informations sur les espaces à risque.

### ***3.2. Le site d'étude : Schwindratzheim, une commune alsacienne fréquemment exposée aux coulées d'eau boueuse***

La commune de Schwindratzheim est une commune rurale de 1574 habitants située dans la plaine d'Alsace (Bas-Rhin) à une vingtaine de kilomètres au NO de Strasbourg (Figure 3).

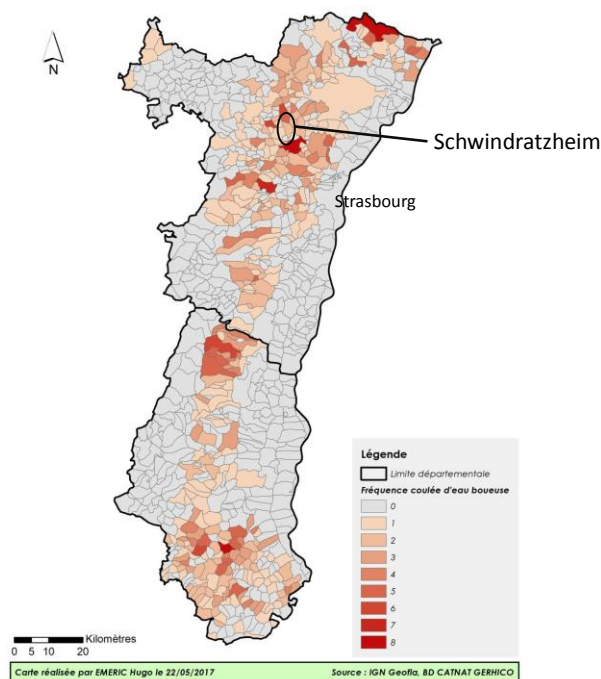


FIGURE 3 : *Fréquence des coulées d'eau boueuse en Alsace et localisation du terrain d'étude.*

Cette commune a été lourdement sinistrée lors d'épisodes de coulées d'eau boueuse (en 1983, 1999 et 2008) après lesquels les élus ont initié de grandes campagnes de prévention du risque. Ainsi, une grande diversité d'aménagements de protection ou de prévention est mise en place dans le paysage agricole : fascines, bandes enherbées, bassins de rétention, assolement concerté, *etc.* (Figure 4). Cette diversité d'aménagements a fortement modifié le paysage de Schwindratzheim, ce qui nous paraît très intéressant dans le cadre de notre enquête.

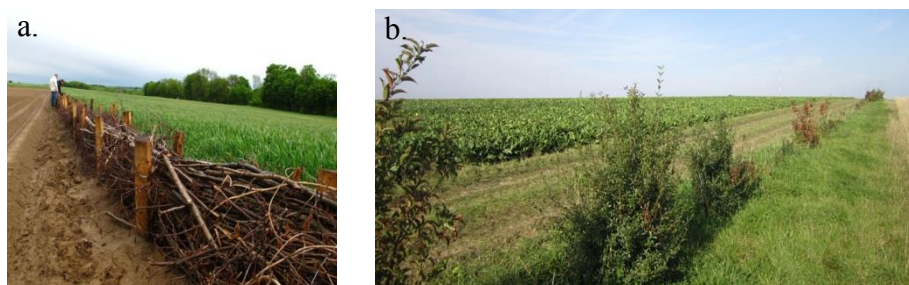


FIGURE 4 : Exemple d'aménagements mis en place à Schwindratzheim (4a : un système de fascines mortes ; 4b : un système de haies).

#### 4. Le traitement des dessins à main levée par un module informatique

Le logiciel que nous avons utilisé a été développé spécifiquement dans le cadre de ce projet. Une première version a été décrite dans Ciron *et al.* (2017). C'est un prototype qui aide à l'analyse de dessins, au moyen de différents algorithmes de traitement d'image. Il a été développé en C++ et est constitué de trois modules : une interface utilisateur, un module de traitement d'image et une base de données.

Une session se déroule en quatre étapes, (1) le chargement des données (dessins et informations sur les personnes enquêtées) via l'interface, (2) la reconnaissance automatique de formes sur les dessins (module de traitement d'image), (3) l'étiquetage interactif des formes extraites, puis (4) la production de synthèses à partir de plusieurs images. Les formes étiquetées sont enregistrées dans la base de données pour un usage ultérieur, dans une visée d'apprentissage automatique.

Plus précisément, le module de traitement d'image opère la détection, la caractérisation et la classification de formes en fonction de critères géométriques en exploitant des règles floues (Fonseca et Jorge, 2000). L'utilisateur peut ensuite étiqueter ces formes selon des types urbains (maison, route, *etc.*) ou liés aux risques (zone de transit, zone d'accumulation). Les formes ainsi étiquetées sont colorées selon le type choisi (Figure 5). L'utilisateur peut également opérer des regroupements (par exemple, si plusieurs maisons sont dessinées à proximité les unes des autres, elles constituent un quartier).

Une fois l'étiquetage terminé, l'utilisateur peut obtenir des statistiques spatialisées selon une grille prédéfinie et sur un ensemble d'images choisies. Ces statistiques prennent deux formes (1) une table, donnant pour chaque cellule de la grille, les pourcentages des différents types d'objets – on peut ainsi voir comment sont localisés les risques pour un groupe d'habitants et (2) une carte moyenne où sont représentés par des niveaux de gris la répétition des mêmes objets dans les différentes images – un bâtiment présent dans la même cellule pour tous les dessins sera affiché en noir, une route représentée seulement dans quelques dessins sera en gris clair.

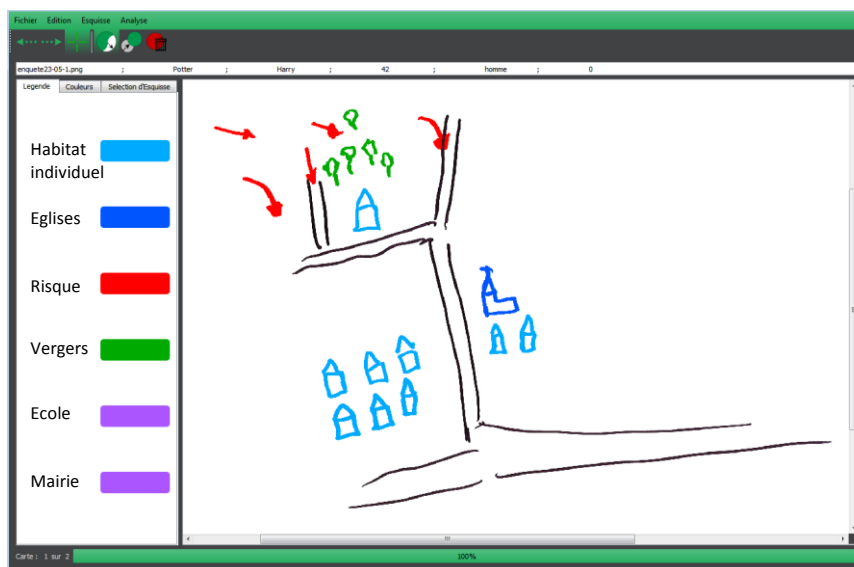


FIGURE 5 : Interface du logiciel, étape (3) : étiquetage interactif des formes détectées - les formes étiquetées apparaissent en couleur selon la légende à gauche.

## 5. Résultats : une étude exploratoire innovante par l'usage d'un traitement informatique des dessins à main levée

Lors des passations effectuées en Juin 2017, nous avons rassemblé un corpus de 20 dessins à main levée (tablettes et au papier-crayon) et 39 questionnaires (certains questionnaires ont été remplis sans être accompagnés d'un dessin pour pouvoir identifier des facteurs sociaux dans les représentations du risque). Etant donné le caractère exploratoire de ce travail, et la non-significativité de notre échantillon, nous ne relèverons ici que quelques tendances à confirmer par des passations supplémentaires sur la commune de Schwindratzheim et/ou sur d'autres territoires.

Notre échantillon se compose de 65% de femmes et 35% d'hommes. 94% des sondés sont propriétaires de leur logement et 80% y vivent depuis plus de 10 ans. 47% de l'échantillon est composé d'un foyer de 4 personnes. Les enquêtés ont un niveau d'étude assez élevé : 50% d'entre eux déclarent avoir un niveau supérieur à Bac + 4-5. Bien que non représentatif de la zone d'enquête, ce premier échantillon permet de mettre en avant certaines tendances que nous exprimons ici.

En reprenant notre hypothèse de départ, à savoir que les représentations socio-spatiales du paysage permettent d'approcher les représentations du risque, nous pouvons mettre en avant trois résultats. Chacun de ces résultats reprend l'un des concepts théoriques retrouvés dans la littérature et présentés ci-avant. Tout d'abord, concernant l'attachement au lieu, les statistiques descriptives (nous utilisons ici l'indice de corrélation linéaire, noté  $r$ ) montrent que moins les enquêtés ont

déménagé au cours de leur vie, plus ils sont attachés à leur lieu de vie ( $r=.40$ ). Il semble également que plus les enquêtés connaissent un nombre important d'aménagements de protection contre les coulées d'eau boueuse, plus ils se sentent attachés à leur lieu de vie ( $r=.35$ ). Et enfin, plus les enquêtés estiment être fortement attachés à leur lieu de vie (« *ma commune est le meilleur endroit pour faire les activités que j'aime* »), plus leur sentiment de protection face à un éventuel risque de coulée d'eau boueuse est important ( $r=.36$ ).

Pour la représentation du risque, les statistiques descriptives montrent clairement que la vue directe sur les aménagements de protection va influencer la représentation du risque, notamment par l'identification de « responsables ». Ainsi, l'analyse de variance (ANOVA) montre que les enquêtés ayant une vue directe sur les aménagements de protection estiment plus que les autres que la taille des parcelles affecte leur exposition au risque ( $F(8,17)$  ;  $p<.01$ ). De même, les enquêtés qui ont une vue directe sur les aménagements sont ceux qui estiment connaître et être suffisamment informés sur les différents types d'aménagements existants ( $F(8,58)$  ;  $p<.01$ ). Nous verrons que cela ne se reflète pas dans la diversité des éléments dessinés.

Enfin, la trajectoire de vie (mesurée via les mobilités tout au long de la vie des enquêtés) influence la connaissance du risque. En effet, plus les enquêtés déclarent avoir déménagé plus de trois fois au cours de leur vie, moins ils semblent connaître le risque de coulées d'eau boueuse ( $r=.27$ ).

Sur les dessins à main levée, en moyenne, les enquêtés dessinent plus d'éléments sur le cadre de vie (avec une moyenne de 15 éléments différents) que sur le risque (en moyenne 5 éléments). Le logiciel permet de spatialiser ces éléments en fonction de la fréquence d'apparition, comme le montre la carte moyenne obtenue à Schwindratzheim (Figure 6).



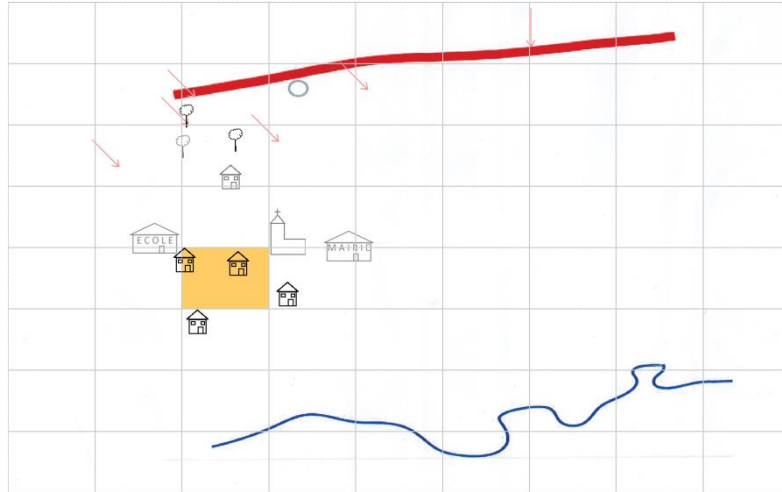


FIGURE 6 : Carte moyenne des éléments du cadre de vie et du risque sur la commune de Schwindratzheim, calculée selon la grille superposée.

Certains éléments sont toujours présents sur le dessin : pour le risque, par exemple, les flèches sont utilisées pour représenter la dynamique et le mouvement de la coulée d'eau boueuse. Elles indiquent toujours la direction préférentielle de cette dernière. En termes d'aménagements présents dans le paysage, seuls les ouvrages hydrauliques (tels que les bassins de rétention) sont représentés sur les dessins. Les aménagements plus petits (dits d'hydraulique douce) n'apparaissent jamais comme étant des éléments structurants.

Pour la représentation du cadre de vie, les habitats sont dessinés sous forme de quartiers, regroupés par zones délimitées (en orange sur la carte moyenne – Figure 6). Une distinction est presque systématiquement faite entre le centre ancien et les nouveaux lotissements. Les nœuds sont les éléments les plus fréquemment dessinés car ils traduisent le réseau de communication du village. Toujours concernant le cadre de vie, les « points de repères » principaux sont les églises, la mairie et l'école.

Un test de Kruskal-Wallis nous permet de vérifier si les dessins diffèrent selon les populations ou les méthodes de production. De fait, nous n'observons aucune différence significative sur le nombre d'éléments entre une passation utilisant le papier-crayon et une passation sur tablette, que cela soit pour la représentation du cadre de vie ( $p=.939$ ) ou du risque ( $p=.878$ ). Il y aurait, en revanche, une différence peu significative en termes de genre ( $p=.051$ ) ou d'appartenance à une classe socioprofessionnelle ( $p=.533$ ) sur la richesse des dessins. Cependant, étant donné la taille réduite de notre échantillon, ces premières analyses doivent être confirmées par davantage de passations.

## Conclusion

À notre hypothèse de départ qui est que les individus entretiennent une relation non intentionnelle au risque auquel ils sont exposés (mesurée via leurs représentations socio-spatiales du paysage), nous pouvons avancer quelques éléments intéressants. Tout d'abord, la littérature montre que les représentations du risque et les représentations socio-spatiales se fondent sur des approches théoriques et méthodologiques proches. Dans ce travail, nous collectons ces données via deux outils : les questionnaires quantitatifs et les dessins à main levée. L'originalité est ici que nous avons créé un logiciel de reconnaissance de formes, d'agrégation de celles-ci et d'analyse. Théoriquement, les bases de notre étude sont posées mais peu de passations ont été effectuées pour l'heure permettant d'asseoir des résultats forts sur l'utilisation conjointe de ces deux outils. Quelques tendances prometteuses ont pu être mises en avant.

Plusieurs perspectives peuvent néanmoins être avancées. Tout d'abord, l'enquête quantitative pourrait mieux préciser les représentations sociales du risque en questionnant davantage les expériences passées, comme par exemple les solutions mises en place à l'échelle individuelle et pouvant contribuer à diminuer la vulnérabilité. Le logiciel pourrait également être amélioré selon différentes voies : amélioration des algorithmes de reconnaissance de forme, étiquetage automatique des objets — en s'appuyant sur la base de données stockant les formes et leurs étiquettes, et ultérieurement, extraction et comparaison de structures spatiales impliquant plusieurs objets d'intérêt comme proposé par Ramaroson *et al.* (2013). De plus, reste en suspens ici l'utilisation pertinente de la 2D pour représenter un phénomène de 3D : notre choix méthodologique se base sur le dessin en vue zénithale, comme cela a été décrit et éprouvé dans la littérature sur les dessins à main levée. Mais cette abstraction introduit-elle un biais cognitif de représentation ? L'utilisation de notre logiciel ne permet pas de répondre directement à cette question. Cependant, elle offre la possibilité de mesurer finement une représentation en vue zénithale, dans l'objectif d'une confrontation avec les résultats d'études utilisant des approches 3D (voir par exemple les travaux de Jacquino et Langumier, 2010).

## Bibliographie

- Appleyard, D., 1970, Styles and methods of structuring a city. *Environment and Behaviour*, 2(1), 100-117.
- Armand, R., 2009, *Étude des états de surface du sol et de leur dynamique pour différentes pratiques de travail du sol. Mise au point d'un indicateur de ruissellement*. Doctorat de Géographie, Université de Strasbourg, 178 p.
- Beck, U., 1985, *La société du risque. Sur la voie d'une autre modernité*, Paris : Aubier.
- Bonaiuto, M., Alves, S., De Dominicis, S., Petrucci, I., 2016. Place attachment and natural hazard risk: research review and agenda. *Journal of environmental psychology*, 48, 33-53.

- Brunet, R., 1974, Analyse des paysages et sémiologie: éléments pour un débat. *L'espace géographique*, 120-126.
- Brunet, S., 2007, *Société du risque, quelles réponses politiques ?*, Paris : L'Harmattan.
- Cadiou, N., Luginbühl Y., 1995, *Modèles paysagers et représentations du paysage en Normandie-Maine*, in : Paysage au pluriel : Pour une approche ethnologique des paysages, Paris : Éditions de la Maison des Sciences de l'homme.
- Caillault, S., Marie, M., 2009, Pratiques agricoles, perceptions et représentations du paysage : Quelles articulations ? Approches croisées Nord/Sud. *Noréis*, 4(213), 9-20.
- Ciron, M., Le Ber, F., Marc-Zwecker, S., Heitz, C., 2017, *Un logiciel pour l'aide à l'analyse de cartes d'acteurs réalisées à main levée*. Atelier Démon associé à la conférence SAGEO, Rouen, 6 novembre 2017.
- Coppieters, Y., Parent, F., Lagasse, R. Piette, D., 2004, Evaluation des risques, une approche pluridisciplinaire en santé publique. *Environnement, Risques et Santé*, 3(1), 45-52.
- Cottet, M., Rivière-Honegger, A., Piégay, H., 2010, Mieux comprendre la perception des paysages de bras morts en vue d'une restauration écologique : Quels sont les liens entre les qualités esthétique et écologique perçues par les acteurs. *Noréis*, 3(216), 85-103.
- Douglas, M., 1992, *Risk and Blame : Essays in Cultural Theory*. London: Routledge.
- Downs, R.-M., Stea, D., 1973, *Cognitive mapping: Images of spacial environments*. Chicago: Aldine.
- Fonseca M. J., Jorge J. A., 2000, *Using fuzzy logic to recognize geometric shapes interactively*. In: Fuzzy Systems, 2000. FUZZ IEEE 2000. The Ninth IEEE International Conference, 1, 291-296..
- Giddens, A., 1991, *A modernity and self-identity*, Standford: Standford University Press.
- Gould, P., White, R., 1984, *Mental maps*. Fribourg: Editions universitaires.
- Heitz, C., 2009, *La perception du risque de coulées boueuses : analyse sociogéographique et apports à l'économie comportementale*. Doctorat de Géographie, Université de Strasbourg, 300 p.
- Hidalgo, M.-C., Hernandez, B., 2001, Place attachment : conceptual and empirical questions. *Journal of environmental psychology*, 21, 273-281.
- Jacopin, R., 2011, *Paysages et pratiques des agriculteurs dans le Sud du Plateau Lorrain : logiques d'organisation et effets sur l'environnement*. Doctorat de Géographie, Université de Nancy 2 – INRA, 591 p.
- Jacquino, F., Langumier, J., 2010, Géovisualisations 3D en action dans l'aménagement du territoire. *Géocarrefour*, 85(4), 303-311.
- Jodelet, D., 1984, *Les représentations socio-spatiales de la ville*, In P.H. Derycke (Ed.), Conception de l'espace, Paris.
- Joffe, H., 2003, Risk: From perception to social representation. *British Journal of Social Psychology*, 42, 55-73.
- Kasperson, R.-E., Renn, O., Slovic, P., Brown, H.-S., Emel, J., Goble, R., Kasperson, J.-X., Ratick, S., 1988, The social amplification of risk: a conceptual framework. *Risk Analysis*, 8(2), 177-187.

- Koh, J., 1988, An ecological aesthetic. *Landscape Journal*, 7, 177-191.
- Krien, N., 2014, *Place des risques côtiers dans la représentation du cadre de vie d'individus possédant des enjeux sur des communes « à risque »*, Doctorat de Psychologie, Université de Bretagne occidentale.
- Le Lay, Y.-F., Piégay, H., 2007, Le bois mort dans les paysages fluviaux français : Eléments pour une gestion renouvelée, *L'Espace Géographique*, 1(36), 51-64.
- Luginbühl, Y., 2008, *Las representaciones del paisaje y sus evoluciones*, in Maderuelo, J. (Ed.), *Paisaje y territorio*, Huesca: CDAN Editores.
- Lynch, K., 1960, *The image of the city*, Cambridge, MA: MIT Press.
- Peretti-Watel, P., 2010, *La société du risque*, Paris : La découverte.
- Piégay, H., Gregory, K. J., Bondarev, V., Chin, A., Dahlstrom, N., Eloise, A., Gregory, S. V., Joshi, V., Mutz, M., Rinaldi, M., Wyzga, B., Zawiejska, J., 2005, Public perception as a barrier to introducing wood in rivers for restoration purposes. *Environmental Management*, 35(5), 665-674.
- Ramadier, T., 2003, *Les représentations cognitives de l'espace: modèles, méthodes et utilité. Espaces de vie: aspects de la relation homme-environnement*, Paris : Armand Colin, 177-200.
- Raymond, C-M., Brown, G., Weber, D., 2010, The measurement of place attachment: personal, community, and environmental connections. *Journal of environmental psychology*, 30, 4, 422-434.
- Ramaroson, J.-H., Le Ber, F., Ramamonjisoa, B., Hervé, D. 2013, Treillis de Galois pour la fusion de connaissances spatiales sur des territoires villageois malgaches. *Revue d'Intelligence Artificielle*, (4-5), 595-617.
- Slovic, P., Peters, E., 2006, Risk perception and affect. *Current directions in psychological science*, 15(6), 322-325.
- Van Dijk, P., Auzet, A.-V., Lemmel, M., 2005, Rapid assessment of field erosion and sediment transport pathways in cultivated catchments after heavy rainfall events. *Earth Surface Processes and Landforms*, 30, 169-182.

---

# Extraction de motifs de trajectoires sémantiques similaires

Clément Moreau<sup>1</sup>, Thomas Devogele<sup>1</sup>, Laurent Etienne<sup>1</sup>

Laboratoire d'Informatique Fondamentale et Appliquée de Tours  
64, avenue Jean Portalis, 37200 Tours France

{clement.moreau,thomas.devogele,laurent.etienne}@univ-tours.fr

---

**RÉSUMÉ.** La compréhension fine des déplacements des individus nécessite une modélisation sémantique riche de leurs activités. Or, il est maintenant possible d'extraire les mobilités et les activités des individus à l'aide d'informations contextuelles ou de capteurs. Une fois enrichies sémantiquement, ces mobilités peuvent être comparées à l'aide d'une ontologie selon une mesure de proximité spatiale, temporelle et sémantique, puis regroupées en clusters de trajectoires similaires. Afin de résumer ces déplacements similaires, des motifs synthétisant ces clusters peuvent être induits. Cet article constitue une étude bibliographique et présente une méthodologie afin d'extraire ces motifs à partir de trajectoires sémantiquement riches. Dans cet objectif, cet article propose de mettre en lumière, d'étendre et de relier un grand nombre d'outils de la fouille des trajectoires. Cette méthode est générique et s'applique à nombre de domaines d'études tels que le tourisme, la sociologie, l'épidémiologie ou l'urbanisme.

**ABSTRACT.** Thanks to the growth of Internet of things and use of various sensors embedded in smartphones, individuals can be tracked and monitored all day long. However, the fine understanding of their activities requires semantic models and contextual information. The users' trajectories can then be enhanced with semantic meaning using an ontology. Trajectories can be compared using their spatial, temporal or semantic components. Similar users' behaviour can then be clustered to derive movement patterns. This article presents a bibliography study and a new methodology about how to extract movement patterns from semantic trajectories. These patterns are helpful for experts working on movement analysis in various fields such as tourism, sociology, epidemiology or urban planning.

**MOTS-CLÉS :** Trajectoires sémantiques, Clusters de trajectoires, Pattern de déplacement, Mesure de similarité, Ontologie, Fouille de trajectoires

**KEYWORDS:** Semantic trajectory, Trajectory clustering, Moving Object patterns, Similarity measure, Ontology, Trajectory datamining

---

## 1. Introduction et contexte

La compréhension de la mobilité humaine est un enjeu dans de nombreux domaines comme l'urbanisme, le tourisme, la sociologie, ou encore l'analyse de propagation des virus. Un résultat primordial donné par (González *et al.*, 2008 ; Song *et al.*, 2010) présente que malgré la complexité et la diversité des trajectoires, la mobilité humaine présente un degré élevé de régularité temporelle et spatiale ce qui indique que les déplacements individuels suivent des modèles reproductibles simples. Cependant, même si les résultats précédemment énoncés mettent en lumière le caractère prédictible des trajectoires humaines, ils ne fournissent pas d'explication satisfaisante quant au sens et au contexte des déplacements (Renso, Trasarti, 2013 ; Yan, 2009 ; Parent *et al.*, 2013).

C'est au sein de cette brèche sémantique que nous posons notre propos. Au cours de notre exposé nous illustrerons notre sujet à l'aide d'exemples tirés de deux projets applicatifs : SMARTLOIRE et MOBI'KIDS, respectivement une plateforme de recommandation pour le tourisme sur mesure en région Centre-Val de Loire, et une étude sociologique visant à comprendre les conditions de mobilités quotidiennes des enfants dans un contexte impulsé par les enjeux de la ville durable et des modes alternatifs de déplacements.

Aussi, l'abondance de ressources complémentaires au GPS, telles que les ontologies ou les médias sociaux tend à affiner la connaissance que nous avons de la mobilité des individus. Dès lors, il convient d'adopter une modélisation des trajectoires en adéquation avec cette richesse sémantique disponible. De cette modélisation découle une métrique permettant de comparer deux trajectoires en vue d'établir un algorithme de partitionnement de données (ou *clustering*) afin de regrouper les motifs (ou *pattern*) de comportements similaires. Une dernière étape se traduit par une représentation synthétique des motifs extraits. L'existence d'une telle chaîne de traitement, analogue à la FIGURE 1. répond entre autres à trois des challenges majeurs proposés par (Ferrero *et al.*, 2016) au sein des SIG actuels, soient : Comment représenter les informations contextuelles et hétérogènes au sein des trajectoires sémantiques (*Multiple Aspect Representation*) ? Comment enrôler les dimensions spatiale, temporelle et sémantique au sein d'une même métrique et établir des partitions de trajectoires proches (*Similarity Analysis and Data Mining*) ? Est-il possible de résumer une partition de trajectoires proches en un motif synthétique (*Vizualisation*) ? Cet article offre un éclairage et des premières réponses à ces problèmes complexes en mettant en lumière un ensemble de méthodes afin de décrire le processus proposé et en adoptant des solutions issues de la littérature. Ainsi, elle se présente avant tout comme un opuscule bibliographique, illustré selon des cas d'école, sur les sujets liés à l'extraction de motifs dans le cadre des trajectoires enrichies sémantiques.

L'article est organisé comme suit : Dans la section 2 , plusieurs modèles de trajectoires sémantiques sont discutés et nous présentons un modèle sémantique adéquat afin de traiter les cas hypothétiques de travail présentés. En section 3, nous passons en revue les différentes mesures de proximité puis présentons une

distance permettant de comparer deux trajectoires au sein de notre modèle, cette distance est reprise en section 4 où sont abordées les problématiques de fouille telles que le partitionnement des trajectoires et l'extraction de motifs fréquents avant de dresser nos conclusions.

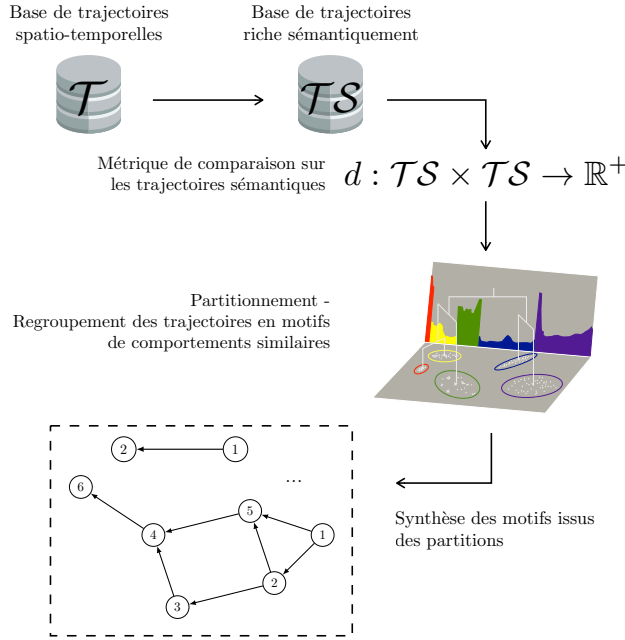


FIGURE 1. Chaîne de traitement pour l'extraction de motifs de trajectoires sémantiques similaires

## 2. Modélisation des trajectoires sémantiques

### 2.1. Les différentes représentations sémantiques des trajectoires

La notion de sémantique au sein des trajectoires spatio-temporelles naquit de la volonté d'introduire une dimension contextuelle au déplacement observé afin de cerner au mieux sa nature. Elle émergea également dans une perspective d'intelligibilité des données afin de réduire le temps de calcul lors des opérations de fouille et d'analyse de données et de rendre l'interrogation plus naturelle. C'est dans cette seconde optique que (Alvares *et al.*, 2007) présente un modèle permettant d'extraire les points de séjour et déplacements (*Stops and Moves*) d'une trajectoire. Les lieux d'arrêt et points de séjour sont ensuite enrichis sémantiquement par un label (hôtel, musée, etc...) en vue d'être analysés de façon

plus haut niveau et non plus seulement géométrique. (Spaccapietra *et al.*, 2008) se réapproprie les notions de Stop and Move en vue d'établir un formalisme de conception de référence pour la modélisation sémantique des trajectoires.

Certains auteurs ont tenté d'élargir la conception de Spaccapietra et al. Par exemple dans (Yan *et al.*, 2010), la trajectoire sémantique est représentée comme une séquence d'épisodes sémantiques et de nombreux modèles comme (Noël *et al.*, 2015 ; Beber *et al.*, 2017) reprennent cette modélisation en adoptant des codes du modèle CONSTAnT établi dans (Bogorny *et al.*, 2014) qui possède un pouvoir d'expression élevé.

Bien souvent la fouille de données exige la prise en compte de paramètres environnementaux (météo, aménagement urbain, lieux d'intérêt) mais aussi thématiques (Transports, horaires d'événements, d'activités, routines ou comportements). C'est notamment en réponse à ces besoins réels que ces modèles plus élaborés, comme CONSTAnT, ont vu le jour. Ils incluent plusieurs notions formalisées telles que les : lieux géographiques et sémantiques, événements, objectifs, activités, environnement et comportement dans l'espoir de devenir une référence dans la conception des trajectoires sémantiques.

Cependant, même si la représentation de Bogorny et al. est particulièrement féconde, elle échoue dans la réponse à l'ensemble des conditions posées par (Yan, 2009) : Une des améliorations significatives serait l'alimentation du contexte des trajectoires par des ontologies afin de mener des inférences sur les activités ou bien d'établir une hiérarchie de concepts qui permettrait alors de calculer plus finement la proximité sémantique de deux entités.

Des vécues énoncées sur CONSTAnT, nous retiendrons le défaut majeur suivant : malgré la volonté de vouloir s'insérer dans une démarche forte pour la fouille de données, le modèle ne fournit malheureusement pas de métrique de comparaison nécessaire. Hormis ce point, nous soutenons la richesse d'expressivité sémantique offerte par le modèle de Borgony et al. et pensons que des modifications telles que celles suggérées dans (Parent *et al.*, 2013 ; Miller, Han, 2009) comme la prise en compte d'ontologies pour l'analyse de données et la généralisation de concepts sont des points-clés à son amélioration.

## 2.2. Modèle de trajectoire riche sémantiquement

Le modèle des trajectoires riches sémantiquement retenu s'appuie sur la notion d'activité. Il reprend les concepts du modèle CONSTAnT et lui associe des informations ontologiques. La FIGURE 2. synthétise ce modèle. Une **trajectoire sémantique** d'individu est une suite d'activités  $TS = \{a_1, \dots, a_n\}$  ordonnée temporellement. Une **activité**  $a$  est décrite selon les trois dimensions : sémantique, spatiale et temporelle. La dimension sémantique est apportée à l'aide d'un ensemble d'ontologies  $\mathcal{O}$ . Les concepts ontologiques sont organisés hiérarchiquement par des liens  $is\_a$  selon un arbre ou un treillis ce qui permet de définir une notion de proximité entre concepts (Aïme, 2011). Par exemple les concepts "Natation" et "Basket-ball" sont des sous-concepts du concept "Acti-



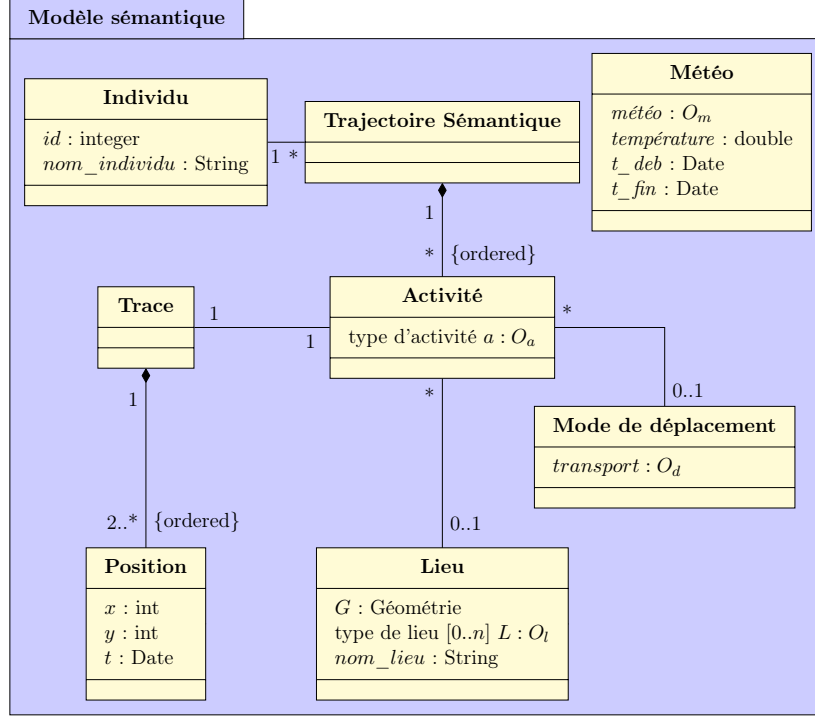


FIGURE 2. Modélisation UML des trajectoires sémantiques d'individus

tivité sportive". Ainsi, dans le cas des activités, tout type d'activité  $a \in O_a$  avec  $O_a \in \mathcal{O}$ , c'est-à-dire que tout concept d'activité doit appartenir à l'ontologie. À noter que le concept `Null` est autorisé, il représente une activité inconnue ou une activité d'attente.

Les activités sont associées à des **traces** [brutes] (ou *raw data*). Ces objets spatio-temporels regroupent des **positions** ordonnées temporellement sous la forme de suites de triplets  $p_i = (x, y, t)$ . Une activité peut-être statique ou mobile, le distinguo est fait si l'on constate que la trace  $T = \{p_1, \dots, p_k\}$  associée à l'activité  $a$  est de la forme  $\exists(x, y) \in \mathbb{R}^2, \forall p \in T, (p.x, p.y) = (x, y)$ , c'est-à-dire que la trace se concentre en un point fixe géographique. De même, c'est la trace qui porte la dimension temporelle et qui délimite la durée d'activité  $d_a$ , soit  $d_a = [p_1.t, p_k.t]$ .

Comme pour le modèle SeMiTri de (Yan *et al.*, 2011), l'alternance entre les activités statiques et mobiles n'est pas obligatoire. Un **mode de déplacement** ("à pied", "à vélo", "en voiture", etc.) est défini également à l'aide d'un concept de mobilité et représente alors une activité mobile.

Un dernier point à signifier est que la trajectoire sémantique est organisée de

manière à éviter tout recouvrement temporel. Ainsi,  $\forall a_i, a_j \in TS, d_{a_i} \cap d_{a_j} = \emptyset$ ; autrement dit, il n'est pas permis d'effectuer plus d'une activité simultanément, ce qui complexifierait le modèle de manière importante; il peut néanmoins être défendu que, dans bien des cas d'activités simultanées, il existe une activité qui prévaut par rapport aux autres. Cependant, si l'on considère la trajectoire sémantique  $TS$  non plus comme plus une suite d'activités ordonnées temporellement mais comme une suite d'ensemble d'activités de la forme  $TS = \{A_1, \dots, A_n\}$  où  $A_i = \{a_1, \dots, a_p\}$ , alors il est possible de tenir compte d'activités simultanées. Cependant, on perd l'ordre temporel total établie entre les activités; les raisonnements temporels peuvent néanmoins être menés l'aide des règles de l'algèbre des intervalles d'Allen.

Les activités sont potentiellement attachées à des **lieux** qui sont des instances du monde physique. Un lieu est décrit par une géométrie  $G$  telle que  $a.T \subseteq G$ , c'est-à-dire que l'on considère le lieu dont la géométrie englobe spatialement l'activité, et un ensemble de concepts de type de lieux  $L \subseteq O_l$  dont les éléments  $l \in L$  sont les parents directs de l'instance décrite. Par exemple, le lieu  $P_1$  est une instance attachée au concept parent  $L_{P_1} = \{\text{piscine}\}$  et  $G_1$  une instance du concept parent  $L_{G_1} = \{\text{gymnase}\}$ . Ces instances sont reliées entre-eux par le concept plus général "équipement Sportif".

En fonction des applications, la météorologie joue un rôle important. Dans ce cas, la **météo** est relatée par une suite de périodes avec deux valeurs : une température et un concept météorologique  $m \in O_m$ .

Les 7 trajectoires schématisques de la FIGURE 3. vont servir d'exemple fil

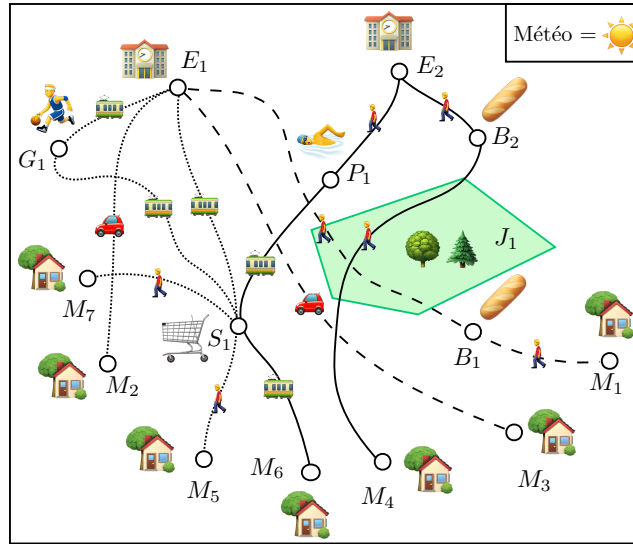


FIGURE 3. Exemple de 7 trajectoires sémantiques d'enfants

rouge à cet exposé. Par convention, la trajectoire sémantique de l'enfant  $i$  se rendant à la maison  $i$  sera appelée trajectoire sémantique  $i$  ( $TS_i$ ). Dans cet exemple, les enfants partent de deux écoles ( $E_1$  et  $E_2$  de concept parent respectif  $L_{E_1} = \{\text{école public}\}$  et  $L_{E_2} = \{\text{école privé}\}$ ) et retournent chez eux à pied, en voiture ou en tramway. On considère que la météo est fixée à "Beau temps" lors de l'analyse.

EXEMPLE. — Modélisation de la trajectoire sémantique  $TS_6$

La trajectoire sémantique  $TS_6$  retrace le retour de l'enfant 6 qui part de l'école  $E_2$  à pied pour se rendre à la piscine  $P_1$  où il nage. Puis, il prend le tramway pour aller faire des courses au supermarché  $S_1$  et finalement rentre en tramway chez lui. Afin de faciliter la compréhension, la dimension temporelle n'a pas été prise en compte dans cet exemple. De même pour les activités statiques, seul le lieu est défini (la trace se réduisant à un point) :  $TS_6 = \langle (\text{apprendre}, E_2), (\text{marcher}, TS_{6.1}, \text{à pied}), (\text{nager}, P_1), (\text{se déplacer}, TS_{6.3}, \text{en tramway}), (\text{faire les courses}, S_1), (\text{se déplacer}, TS_{6.5}, \text{en tramway}), (\text{Null}, M_6) \rangle$ .

Où  $TS_{i.k}$  est la trace de l'activité  $a_k$ , elle n'est indiquée que dans le cadre d'une activité mobile (une activité fixe se réduisant spatialement à un point).  $\square$

Les activités pratiquées à la maison étant inconnues, la dernière activité de cette séquence est Null.

### 3. Mesures de comparaison et paramétrage

#### 3.1. Mesures de proximité classiques

Les contraintes utilisateurs et contextuelles étant des variables imprédictibles, il est nécessaire de constituer une mesure suffisamment adaptative et générique qui puisse correspondre avec précision à l'idée de similarité que porte l'utilisateur ou encore qui lui permette de s'adapter selon ces besoins. Aussi, il subsiste un vide au sein des métriques existantes dans les SIG actuels et qui ne parvient pas à être comblé, une mesure de proximité générique qui puisse faire cohabiter les dimensions temporelle, géométrique et sémantique tout en sachant s'accorder avec les besoins utilisateurs.

Un panorama des mesures de similarité est dressé par (Li, 2014). Du côté géométrique, les distances à représentation linéaire, comme par exemple Dynamic Time Warping (DTW) (Sakoe, Chiba, 1978) et la distance de Fréchet (Devogele, 2002) sont très utilisées dans un contexte d'alignement des données et robustes pour les trajectoires possédant des sinuosités ou boucles. En contrepartie, elles s'appuient d'une matrice de distance d'où une complexité moyenne de calcul en  $O(n \times m)$ , où  $n$  et  $m$  correspondent aux nombres de points des trajectoires. Le pendant sémantique est quant à lui représenté par deux familles de métriques : la distance d'édition et ses variantes - telles que Edit Distance on Real sequence (EDR) (Chen *et al.*, 2005) - basées sur la modification d'une suite de symboles et celles, comme Longest Common Subsequence (LCS), basées sur la recherche de la partie commune contiguë plus longue. Au sujet de la complexité

ces algorithmes, (Wagner, Fisher, 1994) annonce une complexité moyenne en  $O(n \times m)$ . On notera que la distance d'édition est plus adaptative que LCS car elle offre la possibilité de considérer des opérateurs supplémentaires que ceux classiquement proposés (insertion, suppression, modification), de définir les coûts d'opération et tient compte de la totalité de la séquence de trajectoire. Pour ces raisons, nous proposons une métrique sémantique basée sur la distance d'édition et implémentant les opérations de : Ajout, Suppression, Modification ainsi que l'insertion et la permutation. Cette dernière opérations est en accord avec (Furtado *et al.*, 2016) qui avance le fait que deux trajectoires qui visitent les mêmes lieux (même sémantique) mais dans un ordre différent, peuvent être similaires. Les problèmes combinatoires liés à ces opérations de réarrangement ou d'alignement au sein des séquences souvent rencontrés en bio-informatique et commentés chez (Fertin *et al.*, 2009; Marteau, 2009).

Enfin, il est important de préciser que toute entité qualitative doit être considérée au sein d'une ontologie ou hiérarchie de concepts afin de pouvoir être soumise à comparaison. Dans (Aime, 2011) de nombreuses métriques sont proposées afin d'établir la similarité entre deux concepts ce qui permet d'effectuer les opérations de remplacement avec moins de rigidité.

Pour finir, concernant l'agrégation de différentes mesures afin de considérer les aspects géographiques, (Xu, Da, 2003) présente différents opérateurs existants et communément utilisés pour agréger une série de valeurs. Les plus classiques sont les opérateurs min, max, moyenne pondérée; un ensemble des possibilités sur les sommes et moyennes est établi dans (Grabisch *et al.*, 2011). Des solutions pour effectuer la similarité de séquences multi-dimensionnelles sont aussi énoncées dans (Furtado *et al.*, 2016; Gibert *et al.*, 2013).

### 3.2. Mesure générique de similarité entre trajectoires sémantiques

Cette section est dédiée à la présentation de la distance d'édition enrichie permettant d'officialiser sur les trajectoires sémantiques de la section 2. Elle présente entre autres les différents opérateurs considérés puis un exemple d'instanciation de la mesure. L'aspect temporel n'est pas abordé, on suppose néanmoins qu'il est possible d'exercer sur les séquences d'activités un alignement temporel par un algorithme comme DTW.

DÉFINITION 1. — *Opérateurs d'édition*

On définit  $\Sigma = \{(c_a, c) | c_a \in O_a, c \in (O_d \cup O_l)\}$  un alphabet d'activités. Soient deux trajectoires sémantiques  $TS_1, TS_2$  telles que  $TS_1 \in \Sigma^n$  et  $TS_2 \in \Sigma^p$ . On rappelle que  $\varepsilon$  désigne le symbole vide et peut-être rattaché au concept Null. Soient  $TS_1 = \langle a_1, a_2, \dots, a_n \rangle$  et  $TS_2 = \langle a_i, a_j, \dots, a_k \rangle$ . Soient  $a, b \in \Sigma$ , tels que  $a \neq b$  et le couple  $(a, b) \neq (\varepsilon, \varepsilon)$ . On considère l'ensemble d'opérations d'édition  $E = \{\text{add}, \text{supp}, \text{modif}\}$  (ajout, suppression, modification) où  $\forall e \in E, e : \Sigma \times \Sigma \rightarrow \Sigma$ . On définit également l'opérateur d'insertion  $\text{insert} : \Sigma \times \Sigma \rightarrow \Sigma^3$  et l'opérateur sort :  $\Sigma^n \times \Sigma^n \rightarrow \Sigma^n$  de réarrangement.

Ces opérateurs sont définis tels que :

- **modif** :  $a \rightarrow b$  (Remplacement de l'activité  $a$  par  $b$ ).
- **add** :  $\varepsilon \rightarrow b$  (Ajout de l'activité  $b$ ).
- **supp** :  $a \rightarrow \varepsilon$  (Suppression de l'activité  $a$ ).
- **insert** :  $a \rightarrow aba$  (Insertion d'une activité  $b$  durant l'activité  $a$ ). Cet opérateur peut-être utile notamment dans un contexte Move-Stop-Move, pour marquer une course durant un déplacement par exemple. On considérera aussi l'opérateur inverse **insert**<sup>-1</sup> :  $aba \rightarrow a$ .

– On admet que  $n = p$  et que  $\forall a \in TS_1, a \in TS_2$  et  $\forall a \in TS_2, a \in TS_1$ , alors **sort** représente la permutation  $\sigma = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ a_i & a_j & \dots & a_k \end{pmatrix} = \begin{pmatrix} TS_1 \\ TS_2 \end{pmatrix}$ . On représente également  $\sigma$  selon sa décomposition canonique en permutations circulaires telle que  $\sigma = \pi_1 \circ \dots \circ \pi_p$ . On notera  $\lg(\sigma) = p$ .

De plus, on définit la fonction de coût  $\gamma : E \rightarrow \mathbb{R}^+$  qui associe à un opérateur son prix d'utilisation. Il est possible d'ajouter de nouveaux opérateurs, mais il faut veiller à ce que  $\forall e \in E, \exists e^{-1} \in E$  afin de garantir la symétrie.

Une séquence d'opérations d'édition  $(e_1, e_2, \dots, e_N)$  transformant la trajectoire  $TS_1$  en  $TS_2$  est appelée un chemin d'édition  $c(TS_1, TS_2)$  et correspond à la composition successive des opérations  $e_1$  à  $e_N$ . On désigne par  $\mathcal{C}(TS_1, TS_2)$  l'ensemble de tous les chemins d'édition de  $TS_1$  à  $TS_2$ . Le coût d'un chemin d'édition peut alors être déterminé par la somme de ses coûts d'opération d'édition individuels.

DÉFINITION 2. — *Distance d'édition sémantique*

La distance d'édition  $d_S : \Sigma^p \times \Sigma^n \rightarrow \mathbb{R}^+$ , compte tenu de la fonction de coût d'édition  $\gamma$ , est le coût minimal pour transformer  $TS_1$  en  $TS_2$ , soit :

$$d_S(TS_1, TS_2) = \min_{(e_1, \dots, e_N) \in \mathcal{C}(TS_1, TS_2)} \sum_{i=1}^N \gamma(e_i) \quad (1)$$

EXEMPLE. — *Instanciation de  $\gamma$*

Soient les trajectoires sémantiques  $TS_6$  et  $TS_5$  de la FIGURE 3, on détaille  $TS_5$  telle que :  $TS_5 = \langle (\text{apprendre}, E_1), (\text{se déplacer}, TS_{5.1}, \text{en tramway}), (\text{basketball}, G_1), (\text{se déplacer}, TS_{5.3}, \text{en tramway}), (\text{faire les courses}, S_1), (\text{se déplacer}, TS_{5.5}, \text{à pied}), (\text{Null}, M_6) \rangle$ .

On peut définir la fonction de coût  $\gamma$  telle que :

$$\gamma(e) = \begin{cases} 1 & \text{si } e = \text{add ou } e = \text{del} \\ 1 - \text{Sim}(a, b) & \text{si } e = \text{mod} \\ 2\alpha & \text{si } e = \text{insert ou } e = \text{insert}^{-1} \\ \beta \sum_{\pi_i \in \sigma} \sum_{k=1}^{\lg(\pi_i)-1} \text{Sim}(\pi_i(a_k), \pi_i(a_{k+1})) & \text{si } e = \text{sort} \end{cases}$$

où  $\text{Sim} : \Sigma \times \Sigma \rightarrow [0, 1]$  est une fonction calculant la similarité entre un couple d'activités, on notera que  $\text{Sim}$  doit être symétrique, c'est-à-dire que  $\text{Sim}(a, b) = \text{Sim}(b, a)$ <sup>1</sup>; et  $(\alpha, \beta) \in ]0, 1]^2$ , on remarquera que si  $\alpha = 1$  (resp.  $\beta = 1$ ), alors l'opération associée est équivalente à un traitement selon les opérateurs classiques **add**, **supp**, **modif**. Par exemple, pour  $\alpha = 1$ , l'opérateur d'insertion est équivalent à l'application successive de deux ajouts. Une bonne pratique étant de borner le coût des opérateurs entre 0 et 1 pour la modification d'un symbole.

On va transformer  $TS_5$  en  $TS_6$ . On précise que les traces  $TS_{\{5,6\}}.k$  ne sont pas considérées ici mais sont pris en compte dans le calcul d'une distance géométrique  $d_G$ .

On donne :  $\alpha = 0.5, \beta = 0.5$ . Un tel paramétrage permet de réduire de moitié le coût des opérations de permutation et insertion, c'est-à-dire comparative-ment au cas où elles auraient été effectuées à l'aide des opérateurs classiques. On suppose que la fonction de similarité donne les valeurs suivantes pour les couples considérés :  $\text{Sim}((\text{apprendre}, E_2), (\text{apprendre}, E_1)) = 0.85$ ,  $\text{Sim}((\text{se déplacer, en tramway}), (\text{se déplacer, à pied})) = 0.7$ ,  $\text{Sim}((\text{nager}, P_1), (\text{basket-ball}, G_1)) = 0.65$  et  $\text{Sim}((\text{NULL}, M_1), (\text{NULL}, M_2)) = 1$ .

Le chemin d'opérateurs  $c(TS_5, TS_6)$  minimisant  $d_S$  est :

1. **modif**( $TS_5^{(0)}, TS_6^{(0)}$ ) = **modif**((apprendre,  $E_2$ ), (apprendre,  $E_1$ )),
2. **modif**( $TS_5^{(2)}, TS_6^{(2)}$ ) = **modif**((nager,  $P_1$ ), (basket-ball,  $G_1$ )),
3. **sort**( $TS_5, TS_6$ ). On a ici  $\sigma = \pi_1 = (TS_5^{(1)}, TS_5^{(5)})$

L'application de **sort**( $TS_5, TS_6$ ) en toute fin du chemin d'édition permet de permuter les activités  $TS_5^{(1)}$  et  $TS_5^{(5)}$ .

Dès lors  $d_S(TS_5, TS_6) = \gamma(\text{modif}(TS_5^{(0)}, TS_6^{(0)})) + \gamma(\text{modif}(TS_5^{(2)}, TS_6^{(2)})) + \gamma(\text{sort}(TS_5, TS_6)) = (1 - 0.85) + (1 - 0.65) + \frac{1}{2} \times (1 - 0.7) = 0.65$ .

□

Enfin, considérant une distance géométrique  $d_G$  (DTW, Fréchet, ...), la distance  $d_S$  de l'équation (1) et un opérateur d'agrégation  $Agg$ , il est possible de définir notre mesure de proximité  $d : \mathcal{TS} \times \mathcal{TS} \rightarrow \mathbb{R}^+$  telle que :

$$d(TS_1, TS_2) = Agg(d_S(TS_1, TS_2), d_G(TS_1, TS_2)) \quad (2)$$

Supposons alors que  $d_G(TS_1, TS_2) = 3.5$ . On pose l'opérateur d'agrégation moyenne pondérée de dimension 2 :  $Agg(x, y) = \alpha x + (1 - \alpha)y$  avec  $\alpha \in [0, 1]$ . Une telle fonction d'agrégation permet selon le paramètre  $\alpha$  de privilégier la dimension sémantique ou géographique, on posera entre autre  $\alpha = 0.5$  si l'on souhaite traiter ces deux dimension de manière égale.

Supposons que l'on pose  $\alpha = 0.7$  afin de privilégier l'aspect sémantique. Ainsi, l'équation (2) nous donne  $d(TS_1, TS_2) = 0.7 \times 0.65 + 0.3 \times 3.5 = 1.505$ .

---

1. De nombreux exemples de mesures de similarité entre concepts sont exposées dans (Aime, 2011 ; Leacock, Chodorow, 1998 ; Dice, 1945).

## 4. Recherche, partitionnement et synthèse de motifs

### 4.1. Partitionnement des trajectoires sémantiques

Il existe peu de références à notre connaissance sur le partitionnement des trajectoires sémantiques, ceci dû principalement à l'absence d'une métrique pouvant réunir convenablement les dimensions temporelle, spatiale et sémantique. Ainsi, au sein de la fouille des trajectoires, l'aspect par partitionnement fût longtemps envisagé selon le prisme géométrique (Gianotti *et al.*, 2011). Depuis peu, de nouvelles méthodes explorent l'angle sémantique et (Gibert *et al.*, 2013) commente l'apport d'éléments sémantiques et la prise en compte d'ontologies pour les méthodes de clustering hiérarchique (par partitionnement). (Ying *et al.*, 2014) s'approprie les différentes dimensions des trajectoires sémantiques mêlant ainsi partitionnement et fouille de motifs dans un dessein de prédiction. Les auteurs proposent une approche (GTS) basée sur les intentions des utilisateurs selon le contexte géographique, temporel et sémantique pour estimer la probabilité que l'utilisateur visite un lieu. L'idée centrale tient alors dans le calcul d'une similarité entre le mouvement actuel d'un utilisateur et les modèles GTS préalablement découverts.

Cependant, le partitionnement réalisé demeure très dépendant de la modélisation de la trajectoire sémantique et du type de mesure. Par exemple, (Xiao *et al.*, 2014) propose une mesure de similarité sémantique et une approche par partitionnement hiérarchique. Selon les auteurs, deux trajectoires sont considérées comme similaires si elles visitent la même séquence de lieux, plusieurs fois et avec un temps de déplacement similaire. Les permutations sont interdites. Dans une veine similaire que celle réalisée par (Güting *et al.*, 2005), soit selon une méthode de représentation par la description de la position de l'objet par référencement linéaire à l'intérieur d'un réseau d'objets spatiaux en relation, (Wu *et al.*, 2015) propose une métrique selon le triptyque habituel au sein des réseaux routiers en tenant compte également de contraintes temporelles telles que l'horodatage (CTCP).

Enfin, en accord avec les techniques d'extraction d'information issue des médias sociaux et une modélisation de la trajectoire basée sur des régions d'intérêt, (Cai *et al.*, 2016) propose une méthode de partitionnement basée sur la densité. Dans le cadre de la mesure proposée section 3.2, nous soutenons une approche par partitionnement hiérarchique. Si le partitionnement des trajectoires présentées FIGURE 3. est effectué manuellement, deux configurations extrêmes possibles se dégagent : la première est celle où le paramètre géométrique est majoritairement valorisé. Dans ce cas, on observe des partitions figurées par les différents styles de pointillés représentées sur la FIGURE 3., soient  $C_1 = \{TS_1, TS_3\}$ ,  $C_2 = \{TS_4, TS_6\}$  et  $C_3 = \{TS_2, TS_5, TS_7\}$ . Dans l'autre cas, celui où la sémantique prend le dessus, on observe un ensemble de partitions tel que  $C_1 = \{TS_1, TS_4\}$ ,  $C_2 = \{TS_2, TS_3\}$  et  $C_3 = \{TS_5, TS_6, TS_7\}$ .

#### 4.2. Définition de motifs

On a vu en section 2 que de nombreux modèles de représentation de la trajectoire sémantique utilisent des modes de représentation sous forme de séquence. Dans cette lignée, (Gianotti *et al.*, 2007) propose une extension du paradigme de la fouille de motifs séquentiels pour l'analyse des trajectoires d'objets en mouvement. Les auteurs représentent les trajectoires selon le paradigme *Stop and Move* comme des séquences de régions d'intérêt visitées séparées par un temps de déplacement. Le modèle de fouille se base alors sur la découverte de régions d'intérêt en calculant les motifs fréquents de déplacement entre ces régions d'intérêt sous contraintes de seuils spatiaux et temporels. (Zhang *et al.*, 2014) élargit la recherche de motifs de déplacements à des ensembles de points d'intérêt géographiquement compacts, sémantiquement cohérents et dont les transitions temporelles entre ensembles surgissent rapidement (selon un seuil temporel donné). Pour les modèles de mouvement qui sont localement fréquents et non nécessairement dominants dans tout l'espace, (Choi *et al.*, 2017) s'inspire de la notion de compacité utilisée au sein de DBSCAN et l'adapte afin de quantifier la fréquence d'un motif particulier dans l'espace.

Ainsi, la recherche de motifs séquentiels utilise des méthodes à base de seuils (ou supports) qui parfois peuvent manquer de finesse en ne mesurant pas la ressemblance entre deux concepts. Une autre faiblesse est que, dans les modèles présentés, le lieu est considéré indépendamment de l'activité qui peut y être pratiquée; Zhang *et al.*, cependant, argumente le fait que, disposant des informations temporelles et géographiques, des activités peuvent être inférées à partir des médias sociaux. Des exemples allant dans ce sens sont donnés par (Long *et al.*, 2012; Yuan *et al.*, 2012) où les auteurs mettent en vedette une méthode par allocation de Dirichlet latente (LDA) permettant de déduire la fonction d'une région au sein d'une ville (par exemple des lieux d'enseignement, bureaux, zones de commerce) ou encore de déterminer les relations intrinsèques et potentielles entre les lieux géographiques en utilisant les enregistrements de localisation qu'un utilisateur partage sur des médias sociaux. Convaincu d'une influence temporelle forte, (Zion, Lerner, 2017) étend le modèle LDA pour capturer l'influence du temps, en particulier de passé proche (jours/semaines), sur les motifs de mobilité en utilisant des modèles temporels qui assouplissent les hypothèses de LDA afin de considérer au mieux les routines utilisateurs.

Les considérations précédentes montrent la difficulté réelle d'extraire les motifs de déplacement des trajectoires car bien souvent l'information demeure contextuelle. Aussi, si l'on se place dans le cadre d'une méthode par partitionnement préalable, un avantage est que, *a priori*, ces partitions formées offrent des ensembles de trajectoires homogènes aux comportements similaires. Il peut-être souhaitable, par la suite, d'en dresser une synthèse.

Cette vue synthétique peut être de nature géométrique ou sémantique selon les préférences de l'utilisateur. D'un point de vue géométrique, (Etienne *et al.*, 2016) propose le concept de trajectoire médiane sur l'appui de boîtes à moustaches spatio-temporelles. Le pendant sémantique peut quant à lui être



assuré par une représentation sous forme d'automate ou de grammaire formelle (Mouza, Rigaux, 2005). En considérant un alphabet  $\Sigma$  de symboles sémantiques représentant les activités, il est alors possible à l'aide d'une inférence grammaticale sur une partition considérée d'extraire un langage la représentant en substance. Pour l'exemple de la FIGURE 4, cette représentation synthétise un ensemble de déplacements  $\{TS_5, TS_6, TS_7\}$  des enfants qui partent de l'école à pied ou en tramway et qui vont potentiellement faire une activité sportive. Ces enfants prennent ensuite le tramway pour aller faire des courses avant de rentrer à pied ou en tramway chez eux.

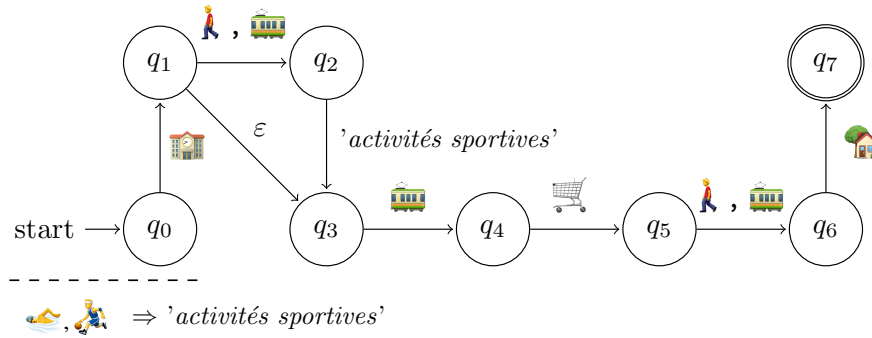


FIGURE 4. Automate et inférence logique représentant le motif synthétique de la partition de trajectoires  $\{TS_5, TS_6, TS_7\}$

## 5. Conclusions

Dans cet article a été présentée une méthode générique pour extraire les motifs de déplacements et d'activités des individus. La connaissance de ces motifs est fondamentale pour mieux comprendre les comportements humains. Ces travaux s'appuient sur une modélisation sémantiquement riche des activités des individus qui intègre les dimensions spatiale, temporelle et sémantique. Pour la dimension sémantique, plusieurs ontologies portant sur les lieux, les activités, les modes de déplacements et la météorologie sont intégrées. De même, ces travaux nécessitent de réutiliser des distances spatiaux-temporelles et d'adapter des mesures de proximité symboliques telles que la distance d'édition. Finalement, ces travaux reprennent les outils de fouille de trajectoires qui extraient premièrement des partitions de trajectoires similaires; pour chaque partition, un motif est inféré dans un second temps synthétisant les trajectoires sémantiques sous la forme d'un automate ou de grammaire. L'idée d'une grammaire probabiliste peut être envisagée afin de refléter le caractère stochastique des déplacements. Un avantage significatif de ces motifs est de résumer de ma-

nière anonyme un ensemble d'activités similaires. Dans le cas d'un nombre de trajectoires suffisant dans chaque groupe, cette synthèse peut répondre à des préoccupations éthiques légitimes pour l'analyse d'activités humaines.

Ainsi cet article, une nouvelle distance d'édition pour la similarité entre trajectoires sémantiques afin d'extraire des motifs de comportement semblables. De futurs travaux doivent être menés afin de rendre cette méthode générique et optimale, en particulier la dimension temporelle doit être approfondie ainsi que les méthodes de similarité entre symboles d'un alphabet afin de rendre complète la métrique, enfin, la distance pourra être mise en œuvre et testée. Également, une solution adaptée pour la gestion simultanée de concepts ontologiques et de données spatio-temporelles volumineuses devra être proposée afin de manipuler ces données hétérogènes de manière optimale.

Cet article forme ainsi un opuscule pour de futurs travaux où sera testée la méthodologie proposée sur des données réelles issues des deux domaines d'application présentés en introduction que sont la mobilité des enfants et les séjours touristiques.

## 6. Remerciements

Ce travail a été financé et soutenu par l'ANR via le projet MOBIKIDS, ainsi que par la région Centre-Val de Loire via le projet de recherche d'intérêt régional SMART LOIRE.

## Bibliographie

- Aime X. (2011). *Gradients de prototypicalité, mesures de similarité et de proximité sémantique : une contribution à l'ingénierie des ontologies*. Thèse de doctorat, Université de Nantes.
- Alvares L., Bogorny V., Kuijpers B., Macedo J. de, Moelans B., Vaisman A. (2007). A model for enriching trajectories with semantic geographical information. *Proc. of the 15th annual ACM international symposium on Advances GIS*, n° 22, p. 1–8.
- Beber M., Ferrero C., Fileto R., Bogorny V. (2017). Individual and group activity recognition in moving object trajectories. *Journal of Information and Data Management*, vol. 8, n° 1, p. 50–66.
- Bogorny V., Renso C., Aquino A. R. de, Lucca Siqueira F. de, Alvares L. (2014). Constant - a conceptual data model for semantic trajectories of moving objects. *Transactions in GIS*, vol. 18, p. 66–88.
- Cai G., Lee K., Lee I. (2016). Discovering common semantic trajectories from geo-tagged social media. In *Trends in applied knowledge-based systems and data science*, p. 320–332. Springer.
- Chen L., Özsu M. T., Oria V. (2005). Robust and fast similarity search for moving object trajectories. *Proc. of the 2005 ACM SIGMOD*, p. 491–502.
- Choi D., Pei J., Heinis T. (2017). Efficient mining of regional movement patterns in semantic trajectories. *Proc. of the VLDB*, vol. 10, p. 2073–2084.

- Devogele T. (2002). A new merging process for data integration based on the discrete fréchet distance. In *Advances in spatial data handling*, p. 167–181. Springer.
- Dice L. (1945). Measures of the amount of ecologic association between species. *Ecology*, vol. 26, p. 297–302.
- Etienne L., Devogele T., Buchin M., McArdle G. (2016). Trajectory box plot; a new pattern to summarize movements. *International Journal of GIS*, vol. 30, p. 835–853.
- Ferrero C., Alvares L., Bogorny V. (2016). Multiple aspect trajectory data analysis: Research challenges and opportunities. *GeoInfo*, p. 56–67.
- Fertin G., Labarre A., Rusu I., Tannier E., Vialette S. (2009). *Combinatorics of genome rearrangements*. The MIT Press.
- Furtado A., Kopanaki D., Alvares L., Bogorny V. (2016). Multidimensional similarity measuring for semantic trajectories. *Transactions in GIS*, vol. 20, p. 280–298.
- Gianotti F., Nanni M., Pedreschi D., Pinelli F. (2007). Trajectory pattern mining. *ACM SIGKDD*, p. 330–339.
- Gianotti F., Nanni M., Pedreschi D., Pinelli F., Rinzivillo S., Trasarti R. (2011). Unveiling the complexity of human mobility by querying and mining massive trajectory data. *The VLDB Journal*, vol. 20, p. 695–719.
- Gibert K., Valls A., Batet M. (2013). Introducing semantic variables in mixed distance measures: Impact on hierarchical clustering. *Knowledge and Information Systems*, vol. 40, p. 559–593.
- González M., CA.Hidalgo, Barabási A.-L. (2008). Understanding individual human mobility patterns. *Nature*, vol. 453, p. 779–782.
- Grabisch M., Marichal J.-L., Mesiar R., Pap E. (2011). Aggregation functions: Means. *Information Sciences*, vol. 181, p. 1–22.
- Güting R., Almeida V. T. de, Ding Z. (2005). Modeling and querying moving objects in networks. *The VLDB Journal*, vol. 15, p. 165–190.
- Leacock C., Chodorow M. (1998). Wordnet: An electronic lexical database. In, p. 265–283. Cambridge MA.
- Li Z. (2014). Spatiotemporal pattern mining: Algorithms and applications. In, p. 283–306. Springer.
- Long X., Lei J., Joshi . (2012). Exploring trajectory-driven local geographic topics in foursquare. *Proc. of the 2012 ACM Conference on Ubiquitous Computing*, p. 927–934.
- Marteau P. (2009). Time warp edit distance with stiffness adjustment for time series matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, p. 306–318.
- Miller H., Han J. (2009). *Geographic data mining and knowledge discovery*. Taylor & Francis.
- Mouza C. du, Rigaux P. (2005). Mobility patterns. *GeoInfo*, vol. 9, p. 297–319.

- Noël D., Villanova-Oliver M., Gensel J., Quéau P. L. (2015). Modeling semantic trajectories including multiple viewpoints and explanatory factors: Application to life trajectories. *Proc. of the 1st International ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics*, p. 107–113.
- Parent C., Spaccapietra S., Renso C., Andrienko G., Bogorny V., Damiani M. *et al.* (2013). Semantic trajectories modeling and analysis. *ACM Computing Surveys*, vol. 45, p. 1–32.
- Renso C., Trasarti R. (2013). Mobility data : Modeling, management and understanding. In, p. 129–151. Cambridge University Press.
- Sakoe H., Chiba S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on ASSP*, vol. 26, p. 43–49.
- Song C., Qu Z., Blumm N., Barabási A.-L. (2010). Limits of predictability in human mobility. *Science*, vol. 327, p. 1018–1021.
- Spaccapietra S., Parent C., Damiani M., Macedo J. de, Porto F., Vangenot C. (2008). A conceptual view on trajectories. *Data & Knowledge Engineering*, p. 126–146.
- Wagner R., Fisher M. (1994). The string-to-string correction problem. *Journal of the ACM*, vol. 21, p. 168–173.
- Wu X., Zhu Y., Xiong S., Peng Y., Peng Z. (2015). A new similarity measure between semantic trajectories based on road networks. *Proc. of the 17th Asia- Pacific Web Conference*, p. 522–535.
- Xiao X., Zheng Y., Luo Q., Xie X. (2014). Inferring social ties between users with human location history. *Journal of AIHC*, vol. 5, p. 3–19.
- Xu Z., Da Q. (2003). An overview of operators for aggregating information. *International Journal of intelligent systems*, vol. 18, p. 953–969.
- Yan Z. (2009). Towards semantic trajectory data analysis: A conceptual and computational approach. *VLDB Endowment*.
- Yan Z., Chakraborty D., Parent C., Spaccapietra S., Aberer K. (2011). Semitri: A framework for semantic annotation of heterogeneous trajectories. *Proc. of the 14th International Conference on Extending Database Technology*, p. 259–270.
- Yan Z., Parent C., Spaccapietra S., Chakraborty D. (2010). A hybrid model and computing platform for spatio-semantic trajectories. *Proc. of the 7th international conference on The Semantic Web*, p. 60–75.
- Ying J.-C., Lee W.-C., Weng T.-C., Tseng V. (2014). Mining geographic-temporal-semantic patterns in trajectories for location prediction. *ACM Transactions on Intelligent Systems and Technology*, vol. 5, n° 2, p. 1–33.
- Yuan J., Yu Z., Xing X. (2012). Discovering regions of different functions in a city using human mobility and pois. *Proc. of the 18th ACM SIGKDD*, p. 186–194.
- Zhang C., Han J., Shou L., Lu J., Porta T. L. (2014). Splitter: Mining fine-grained sequential patterns in semantic trajectories. *VLDB Endowment*, p. 769–780.
- Zion E., Lerner B. (2017). Learning human behaviors and lifestyle by capturing temporal relations in mobility patterns. *Proc., European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, p. 459–464.

---

# Analyse d'image par réseaux de neurones pour les malvoyants

Said Boularouk<sup>1</sup>, Didier Josselin<sup>2</sup>, Eitan Altman<sup>3</sup>

1. LIA - Université d'Avignon

said.boularouk@univ-Avignon.fr

2. UMR 7300 ESPACE

didier.josselin@univ-avignon.fr

3. NEO INRIA Sophia Antipolis

eitan.altman@inria.fr

---

*RÉSUMÉ.* Nous proposons ici une application capable d'analyser une image, d'en extraire les objets en fonction d'une légende et ainsi de décrire une cartographie en ligne. Nous utilisons des réseaux de neurones (perceptrons multi-couches à mémoire) en série ou en parallèle, dont la fonction est d'apprendre à extraire les motifs de l'image à partir de ses couleurs, à différentes échelles. Cette application est développée pour aider à terme les personnes malvoyantes à se représenter l'espace proche de leur localisation pour mieux s'y mouvoir.

*ABSTRACT.* In this paper, we propose an application able to analyse an image, to extract its objects according to the legend and globally to describe an online map. We use artificial neural networks (multilayer perceptrons with memory) in series and in parallel, whose objective is to learn the image patterns according to colors at different scales. This application is designed to help blind people to represent the surrounding space in order to better move on.

*MOTS-CLÉS :* malvoyants, cartographie, raster, réseaux de neurones, texte, intelligence artificielle

*KEYWORDS:* blind people, maps, raster, neural network, text, artificial intelligence

---

## 1. Introduction

Pour parvenir à comprendre et à maîtriser l'espace, l'homme a collecté des données géographiques pour les étudier, les analyser et les partager avec sa communauté. Avec l'émergence de bases de données spatiales, Les SIG ont donné la possibilité de stocker et d'interroger les caractéristiques spatiales d'un objet (Zee, Scholten, 2013). La modélisation spatiale, processus d'abstraction de l'espace réel (espace vécu, perçu, représenté), permet de fournir une représentation simplifiée de l'espace géographique. Celle-ci peut classiquement prendre la forme de représentations vectorielles ou matricielles (Morizot, 2010) et s'appuyer sur des ontologies, comme nous l'avons montré dans un précédent article sur la transcription verbale des données vectorielles pour les malvoyants à partir de leurs attributs (Boularouk *et al.*, 2016). Dans la continuité de ce travail, nous proposons une méthode complémentaire, qui vise à extraire une information spatiale à partir d'images. Il s'agit en quelque sorte d'une transcription langagière de l'espace à partir d'une image raster.

Dans les années 90, les travaux sur l'intelligence artificiels et notamment les réseaux neuromimétiques (Jodouin, 1994) se sont développés à grande vitesse, tant des points de vue théorique (Sarzeaud, 1995) que mathématique et appliquée (Abdi, 1994). En géographie, une première thèse a été publiée sur les réseaux de neurones et les systèmes d'induction appliqués à la prévision de la déprise agricole en 1995 (Josselin, 1995). En 1997, Stan et Christine Openshaw ont publié le premier ouvrage sur l'intelligence artificielles en géographie (Openshaw, Openshaw, 1997). Puis, les travaux dans ce domaine se sont mis quelque peu en sommeil.

La classification d'image est redevenue depuis quelques années l'un des sujets les plus importants de recherche dans les domaines de la vision par ordinateur, la reconnaissance d'objets et le *machine learning* (Heaton, 2015), domaines dans lesquels les réseaux de neurones artificiels sont particulièrement développés (Bishop, 2006). Ils ont atteint des performances exceptionnelles dans la vision par ordinateur, telle que la reconnaissance d'objets sur des images. Un exemple original est celui de la détection de déversements d'hydrocarbures à partir d'images radars de la surface de la mer analysés via des images par des drones dotés de logiciels de machine learning open source de google - *TensorFlow*<sup>1</sup>) (Kubat *et al.*, 1998 ; Abadi *et al.*, 2016).

La reconnaissance des images est souvent réalisée à l'aide d'une procédure de classification, qui peut être divisée en trois grandes étapes (Doan, Poulet, 2014) : extraction des linéaments, encodage des linéaments et apprentissage. Nous proposons ici une méthode de reconnaissance et de transcription en texte des portions de l'espace géographique considérés comme des "motifs spatiaux". La cartographie en ligne est un ensemble d'images spatiales, considérée comme

---

1. <https://www.tensorflow.org/>

une source et un support de stockage riche en informations facilement accessible. Pour en faire bénéficier les personnes aveugles, il serait intéressant de proposer un narrateur capable de décrire les représentations graphiques de l'espace durant la navigation sur une carte. Cet article propose une méthodologie qui pourrait à terme servir dans la détection dynamique d'objets sur une image ou une carte et leur transcription en texte, via des matériels embarqués couramment utilisés par les malvoyants et gérant des fonctions de *Text-to-Speech*. Ne sont développées ici que les méthodes d'apprentissage et d'identification des objets par l'usage des réseaux de neurones.

## 2. Acquisition de la connaissance spatiale chez les personnes malvoyantes

### 2.1. Carte mentale et description langagière de l'espace

La carte mentale est une représentation cognitive subjective de l'espace dont l'objectif est de conceptualiser l'espace. Dans le cas d'une personne déficiente visuellement, il était important de savoir si la transcription de l'information spatiale par un champ autre que la vision est pertinente ou pas, afin que les aveugles créent leurs propres images mentales de l'espace à partir d'une information basée sur des sons (langagière ou musicale).

Une étude menée par le laboratoire LIMSI - CNRS de Paris (France) sur la construction des images mentales à partir d'une description verbale d'objets, et non pas à partir d'une expérience visuelle, a montré que les individus sont capables de construire des images mentales à partir de descriptions verbales (Denis, Gonçalves, 1995). Ces images ont des propriétés similaires à celles des images dérivées de la perception. La similarité est telle que les représentations mentales ont le même contenu et que les fonctions cognitives sont aussi similaires. Les réponses sur des questions, concernant la comparaison des distances après mémorisation d'une carte, montrent une corrélation avec les réponses des personnes qui ont reçu la description verbale de la carte. Les expériences ont montré une augmentation significative du flux sanguin dans le cortex temporo-occipital, quand les participants repensent ou scannent mentalement la carte. Ce résultat est compatible avec l'idée que les représentations visuelles construites à partir de l'expérience visuelle et celles construites à partir de description verbale impliquent les mêmes zones corticales (Denis, Gonçalves, 1995) et que les personnes malvoyantes développeraient des capacités équivalentes de représentation spatiale (Josselin *et al.*, 2016).

Une autre étude réalisée au département de psychologie et de neurosciences de l'université Johns Hopkins à Baltimore (USA), montre que les régions occipitales répondent à la langue parlée dès la petite enfance. À quatre ans, le cortex occipital latéral répond plus aux histoires parlées qu'à la parole dans une langue étrangère ou à la musique. Bien que les réponses langagières soient présentes au début, les réponses occipitales à tous les sons continuent d'augmenter

entre quatre et dix-sept ans (Bedny, Richardson, 2015). Ainsi, chez les enfants malvoyants, une description langagière en langue maternelle d'un lieu spatial serait plus efficace qu'une description musicale, déjà développée dans une publication précédente (Josselin *et al.*, 2016). Ces résultats militent en faveur d'une transformation de l'information spatiale transmise par la vision en une information sous forme verbale et donc, pour ce qui concerne les malvoyants, du développement d'outils utilisant ce vecteur.

## 2.2. Synthétiseurs vocaux d'information en ligne

La synthèse vocale est une technique bien connue et largement utilisée par les personnes malvoyantes, sous forme de synthétiseurs vocaux pour écouter les contenus numériques, agissant comme des substituts de voix et fournissant aussi un outil de compensation pour les personnes qui ne peuvent pas communiquer verbalement. Les lecteurs d'écran ou narrateurs sont des synthétiseurs vocaux qui traduisent le contenu de l'écran en texte parlé. Ils ont été développés comme une fonction d'accessibilité aux personnes malvoyantes et aveugles sur l'ordinateur dans beaucoup de marques du marché : Nextup, Wordtalk, NVDA, Naturalreader, EText Reader IGT, Balabolka Spokentext, Jaws, Narrateur Windows 10, Supernova, Voiceover... Certains lecteurs d'écran sont intégrés dans les navigateurs web (ChromeVox, Talkback, Voiceover). Ces lecteurs ne peuvent décrire une image ou une cartographie que si celle-ci contient une description verbale. C'est dire si l'enjeu de reconnaissance des objets et de leur traduction en texte est important et novateur.

Dans ce contexte, le W3C World Wide Web Consortium a créé le groupe WAI (Web Accessibility Initiative) pour veiller à l'accessibilité des personnes en situation de handicap au contenu du web. Le WAI lance périodiquement des directives appelées WCAG (Web Content Accessibility Guidelines) aux créateurs des sites web et aux développeurs des outils de création du contenu. Les objectifs de ces directives sont de rendre les pages du web<sup>2</sup> :

- *perceptives* : fournir des alternatives textuelles à tous les contenus non textuels (image, représentation graphique...);
- *utilisables* : garantir aux utilisateurs handicapés un temps suffisant pour comprendre et utiliser le contenu;
- *compréhensibles* : fournir des textes compréhensibles;
- *robustes* : optimiser la compatibilité avec les utilisateurs actuels et futurs, y compris avec les technologies d'assistance.

En France, les règles d'accessibilité web aux personnes handicapées sont gérées par le RGAA (Référentiel Général d'Accessibilité pour les Administra-

---

2. <http://www.certam-avh.com/>



tions<sup>3</sup>), qui stipule que les sites web doivent contenir certaines propriétés particulières afin qu'il soient consultable par tous, sans discrimination d'accès.

### 2.3. Un SIG adapté aux malvoyants

Quand la personne aveugle ou malvoyante souhaite lire *auditivement* une portion de l'espace choisi, elle clique sur la cartographie (exemple : Google Maps, OpenStreetMap, etc). Le système capture la représentation spatiale sélectionnée et la convertit en image. Le nombre de couleurs dans cette image est beaucoup plus grand que le nombre de couleurs définies par la légende de la cartographie (cf. Figure 1<sup>4</sup>). L'idée donc est de quantifier les couleurs de l'image capturée et, à partir de cette quantification, donner à la personne aveugle une description de la représentation spatiale choisie en se basant sur les couleurs figurantes sur cette dernière.

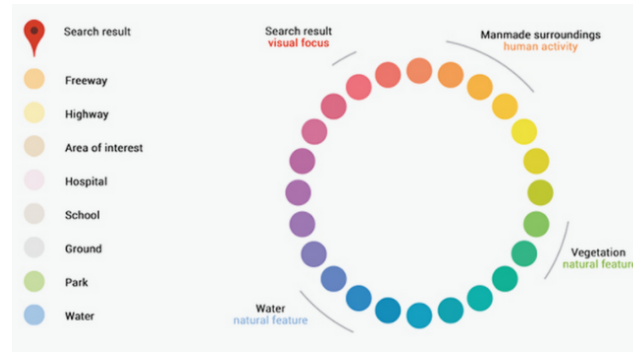


FIGURE 1. Légende de Google Maps

## 3. Quantification des couleurs dans une image

### 3.1. Objectif de la quantification

La quantification des couleurs est le fait de réduire le nombre de couleurs dans une image. une quantification peut être traitée comme un problème de segmentation d'image. Un exemple de quantification est présentée pour l'obtention de cartes sonores pour les aveugles (Josselin *et al.*, 2016). Il semble qu'il n'y ait pas encore de théorie universelle sur la segmentation des images en couleurs (Cheng *et al.*, 2001).

La figure 2 représente la distribution des couleurs dans Google Maps, en RGB. Nous comptons 19 couleurs, chaque couleur représentant une compo-

3. <https://references.modernisation.gouv.fr/rgaa-accessibilite/>

4. <https://blog.google/products/maps/discover-action-around-you-with-updated>

sante spatiale (lieu, route, etc). La figure 3 représente la distribution des couleurs de l'image capturée de la carte Google Maps. On compte ici 450 couleurs différentes. La première étape consiste donc à classifier ces couleurs suivant les couleurs de la légende (les couleurs de construction de la carte).

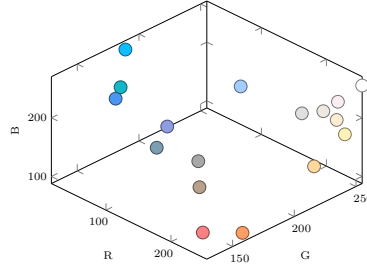


FIGURE 2. Nombre de couleurs définie par la légende Google Maps

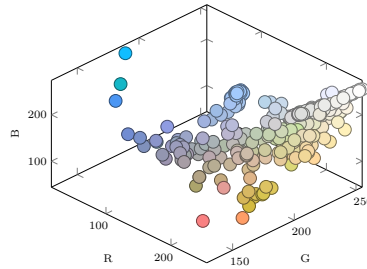


FIGURE 3. Nombre de couleurs différentes dans une image capturée de la carte Google Maps

### 3.2. Méthodes de quantification des couleurs

Différentes méthodes de quantification des couleurs existent. L'algorithme du *K-means* est l'une des méthodes les plus simples et les fréquemment plus utilisées en segmentation pour trouver les centroïdes des segments dont le nombre est défini préalablement par l'utilisateur (Chen *et al.*, 2008). La méthode du *K-means* se base sur la réduction de la distance euclidienne, dans l'espace *RGB* ou *HSV*, entre les centroïdes (choisis aléatoirement) et les pixels de l'image. Le *K-means* est considéré comme exigeant en temps de calcul dans la segmentation des images en couleurs (Aqil Burney, Humera, 2014). Le principe du *k-means* est de minimiser l'erreur quadratique entre les points d'un cluster et son centroïde (Celebi, 2011).

Il est aussi possible de quantifier les couleurs de l'image par l'usage de réseaux de neurones artificiels. Dans ce cas, le problème est de réaliser une classification des couleurs de l'image en fonction des couleurs de la légende. Par exemple, un réseau de neurones de Hopfield constitue une méthode intéressante

de segmentation d'image en couleurs, appliquée dans l'imagerie satellitaire, par exemple en apiculture (Sammouda *et al.*, 2013). Nous utilisons pour notre part un perceptron classique à mémoire (conservation des poids de la couche cachée) à trois couches, dont une couche cachée. En effet, cette structure permet un apprentissage non linéaire (Abdi, 1994). Nous dessinons le réseau de neurones que nous avons utilisé dans la figure 4.

La fonction d'activation (dite de transfert) est la fonction sigmoïde (Jodouin, 1994) (Figure : 5) :  $\sigma(x) = \frac{1}{1+\exp(-x)}$  avec  $(\sigma(x) \rightarrow 1, x \rightarrow +\infty)$  et  $(\sigma(x) \rightarrow 0, x \rightarrow -\infty)$

La fonction dérivée de  $\sigma(x)$  est :  $\sigma(x) * (1 - \sigma(x))$

Les entrées de ce premier réseau de neurones (Fig :4) sont les pixels de l'image capturée  $P$  en format RGB.

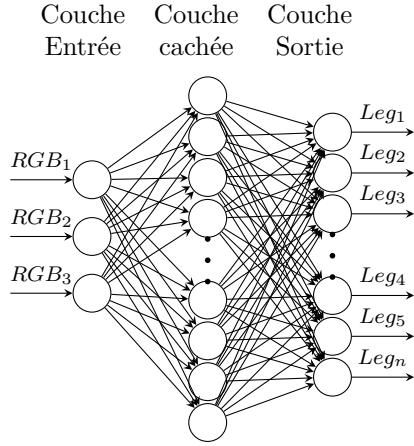


FIGURE 4. Réseau de neurones

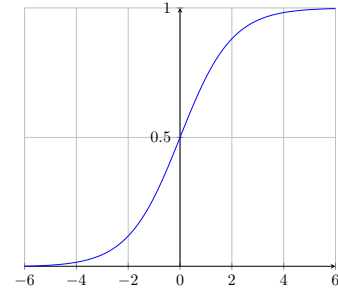


FIGURE 5. Fonction Sigmoïde

Les neurones de la couche d'entrée sont  $e_1, e_2, \dots, e_n$ , chacun prend une valeur  $[R, G, B]_n$ . Les neurones de la couche cachée sont  $h_1, h_2, \dots, h_n$ . Les neurones de la couche de sortie sont  $s_1, s_2, \dots, s_n$ . Ces derniers représentent un vecteur dont les valeurs sont 0 ou 1. Sa longueur est le nombre de couleurs de la légende. Chaque neurone représente une couleur de la légende  $Leg_n$ . Les poids synaptiques entre la couche d'entrée et la couche cachée sont  $w_1, w_2, \dots, w_n$ . Les poids synaptiques entre la couche cachée et la couche de sorties sont  $w'_1, w'_2, \dots, w'_n$ .

Pour l'apprentissage, la fonction de transfert est définie par :

$$\sigma(e) = \frac{1}{1+\exp(-\sum_i^n w_i * e_i)}.$$

Dans cette phase de quantification, les valeurs des pixels sont modifiées par la formule  $(220 - P)/255$   $P \in [R, G, B]$ , nécessaire pour différencier les niveaux

de gris (par exemple les bâtiments sont codés  $[242, 242, 242]$ ) avec un minimum d'itérations.

#### 4. Classification de l'image colorée à multiple résolution

Pour mieux comprendre et interpréter l'espace, nous prenons en considération l'échelle de la carte, matérialisée en quelque sorte par son niveau de zoom et son niveau de détails. Comme nous allons le voir, cela permet de mieux discriminer certaines composantes spatiales et mieux affiner la description de l'espace en fonction de la granularité spatiale des objets à détecter dans l'image. Plusieurs quantifications sont ainsi réalisées à différentes échelles pour le territoire considéré.

Après la phase d'apprentissage, et ce à un niveau d'échelle donnée, le réseau de neurones est capable de classifier toutes les couleurs de l'image suivant les couleurs de la légende. Cette quantification nous permet de calculer facilement la quantité des couleurs correspondant à celles de la légende et aussi les couleurs dominantes dans l'image.

Les valeurs des poids synaptiques sont stockées dans des fichiers pour leur utilisation ultérieure. Elles facilitent ensuite la quantification des couleurs de chaque portion de l'espace choisie par l'utilisateur (phase de reconnaissance). La figure 6 représente les résultats de quantification par le réseau de neurones. Les images sont fractionnées respectivement en  $28 \times 28$  et  $64 \times 64$  pour les deux imageries à droite de la figure.



FIGURE 6. Exemple de segmentation  $28 \times 28$  et  $64 \times 64$  sur la ville d'Avignon

Avec Le *K-means* on traite le problème de quantification de l'image comme un problème de segmentation en un nombre fini de clusters (couleurs de la légende), les valeurs des centroïdes changent de valeurs tout le temps. Pour cela on effectue une classification de chacun des centroïdes obtenus à une classe de couleurs de la légende dont la distance euclidienne (1) est la plus proche. Les réseaux de neurones traitent la quantification directement comme un problème de classification.

$$\sqrt[2]{(R_2 - R_1)^2 + (G_2 - G_1)^2 + (B_2 - B_1)^2} \quad (1)$$

avec  $(R_1, G_1, B_1)$  sont les valeurs du centroïde trouvé et  $(R_2, G_2, B_2)$  sont les valeurs de chaque couleur de la légende.

Les deux méthodes précédentes permettent une quantification des couleurs du portion de l'espace choisi. Par contre, les réseaux de neurones nous semblent plus concluantes en ce qui concerne la qualité de l'image quantifiée, et plus économique en ressource informatique, puisque la même classe des réseaux de neurones suffira pour traiter l'étape de reconnaissance qu'on va développer dans ce qui va suivre. Nous allons par la suite procéder à la transcription vocale de cet espace, pour cela on a besoin de connaître ses composantes.

## 5. Reconnaissance des "motifs spatiaux"

Les "motifs spatiaux", tels que nous les entendons, constituent des objets (agrégats de pixels) dont le réseau de neurones permet de capturer dans l'espace :

- les éléments significatifs,
- les relations existantes entre ces éléments,
- les propriétés des éléments et de leurs relations,

pour leur affecter un nom, un type, des coordonnées géographiques.

Le processus de reconnaissance proposé dans cet article vise l'extraction des éléments importants représentés dans l'image, sans utiliser les métadonnées stockées sous formes textuelles dans les bases des données vectorielles, comme on l'a déjà présenté dans l'article (Boularouk *et al.*, 2016), mais directement à partir de la composition et de l'organisation locales des pixels dans la représentation graphique.

### 5.1. Reconnaissance des compositions colorées locales par des réseaux de neurones en série

La reconnaissance des motifs spatiaux est complétée par un autre réseau de neurones placé en aval du premier. Le vecteur d'entrée du second réseau est l'ensemble des quantités des couleurs résultantes du premier réseau. Le vecteur de sortie est l'ensemble des différentes descriptions sémantiques que l'on donne à l'espace.

Pour les personnes dotées de vision, la reconnaissance des motifs spatiaux sur une carte dépend de la perception visuelle de l'image visualisée, par rapport à l'échelle (zoom), à la légende descriptive et suivant la sémiologie graphique (Bertin, 1973) régissant ladite cartographie (formes, couleurs). L'association est réalisée en fonction de l'expérience du sujet qui l'amène à catégoriser l'ensemble des objets qu'il observe sur la carte. Dans notre cas, il s'agit de mettre en correspondance les compositions des couleurs obtenues par le premier réseau

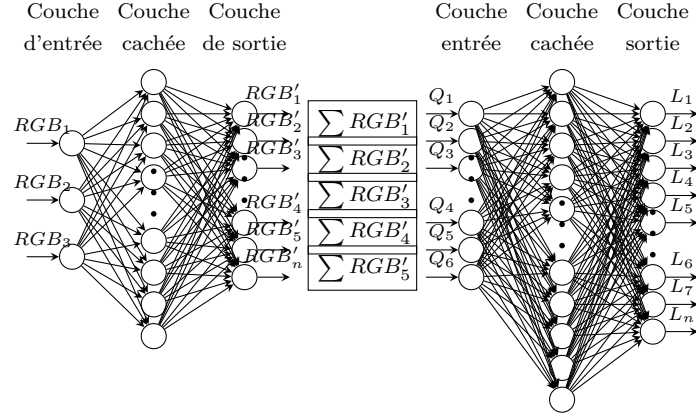


FIGURE 7. Réseaux de neurones en série pour la reconnaissance des portions de l'espace

de neurones avec des catégories prédéfinies d'occupation et d'utilisation du sol, telles qu'on les trouve dans les cartes. Ainsi, le vecteur de sortie regroupe les noms des différentes composantes spatiales et urbaines souvent représentées sur une cartographie. On trouve par exemple : "espace vert", "fleuve", "zone d'intérêt", "forêt", "océan ou mer", "étang d'eau", "réseau routier", "bâtiments", "rue", "autoroute", "hôpital", "intersection rue", "intersection autoroute".

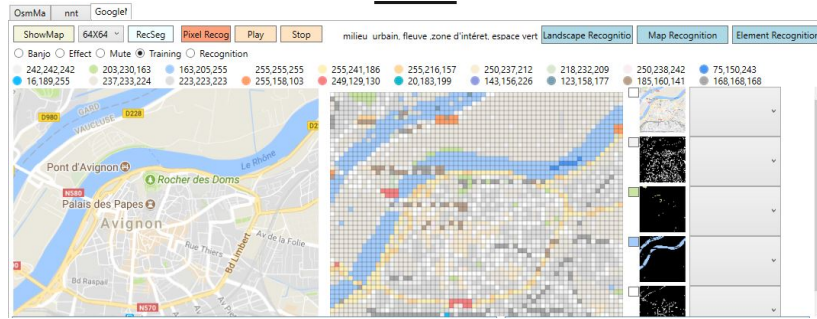


FIGURE 8. Exemple de reconnaissance de motif spatial : "milieu urbain, fleuve, zone d'intérêt, espace vert".

Un motif spatial, urbain ou non, peut être un ensemble des composantes précédentes. Les combinaisons des noms des composantes précédemment citées deviennent aussi des éléments du vecteur de sortie, par combinaison (par exemple : "fleuve, espace vert, zone d'intérêt, bâtiments"). Après une phase d'apprentissage, le système est capable de retrouver les éléments (d'après leurs couleurs) qui composent la portion de la cartographie choisie, comme le montrent la figure 8 pour l'association *milieu urbain, fleuve, zone d'intérêt, espace vert* et la figure 9 pour l'association *milieu urbain, zone d'intérêt*. C'est cette association

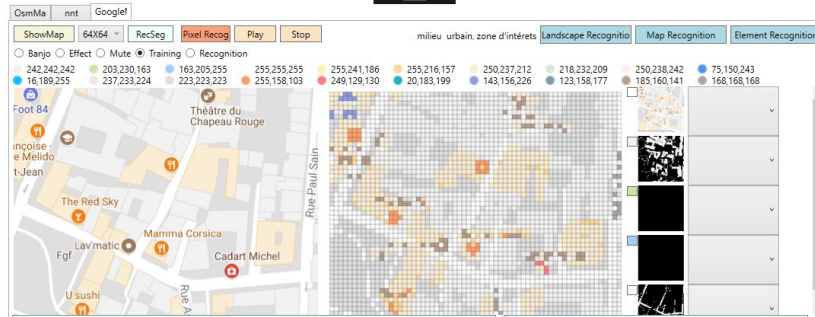


FIGURE 9. Exemple de reconnaissance de motif spatial : "milieu urbain, zone d'intérêt"

des couleurs décrivant la carte qui constitue, selon nous, l'apport essentiel du réseau de neurones lors de son "apprentissage" des objets de l'image.

Cette description peut être complétée par une reconnaissance de l'image décomposée en plusieurs couches.

## 5.2. Reconnaissance des formes et d'organisations spatiales dans l'image par des réseaux de neurones en parallèle

La quantification permet de décomposer la représentation graphique de l'espace en une représentation en couches séparées où chaque couche représente une catégorie de la composante spatiale. La figure 10 présente un exemple de vue de la ville de Marseille sur Google Maps, quantifiée avec une résolution de  $28 \times 28$ . Les imagerie avec un fond noir sont les couches extraites. Chacune d'entre elles a une couleur qui la représente.

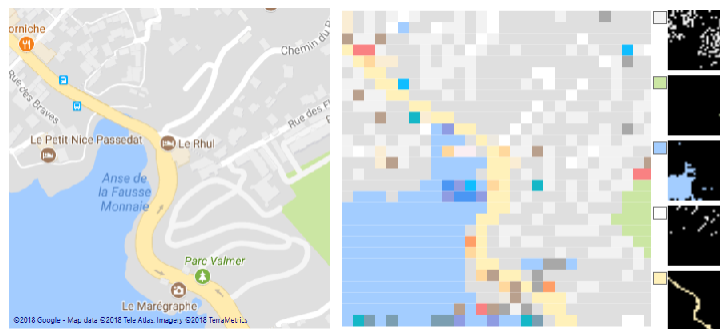


FIGURE 10. Extraction des couches

L'idée est ici que chaque couche soit numérisée de façon binaire (le noir par 0 et la couleur par 1) et devienne l'entrée d'un réseau de neurones indépendant

(cf. Table 1). L'apprentissage se fait dans le but d'extraire des informations plus précises sur l'espace, c'est-à-dire :

- les formes des routes, des rues, des autoroutes, des fleuves,
- la densité des forêts, des zones d'intérêts,
- l'orientation des bâtiments,
- la position de la composante spatiale dans l'image (en haut, en bas, à gauche, etc.).











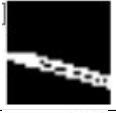




Couleur bleue	Couleur blanche	Couleur verte
 Fleuve	 corner	 Forest(16)*
 Déviation en Y	 Giratoire	 Forest(9)
 cote maritime	 Intersection	 Jardin(16)
 Plusieurs lacs	 est-ouest	 Stade(18)
 Lac	 Nord-Sud	 Plusieurs Jardin(15)

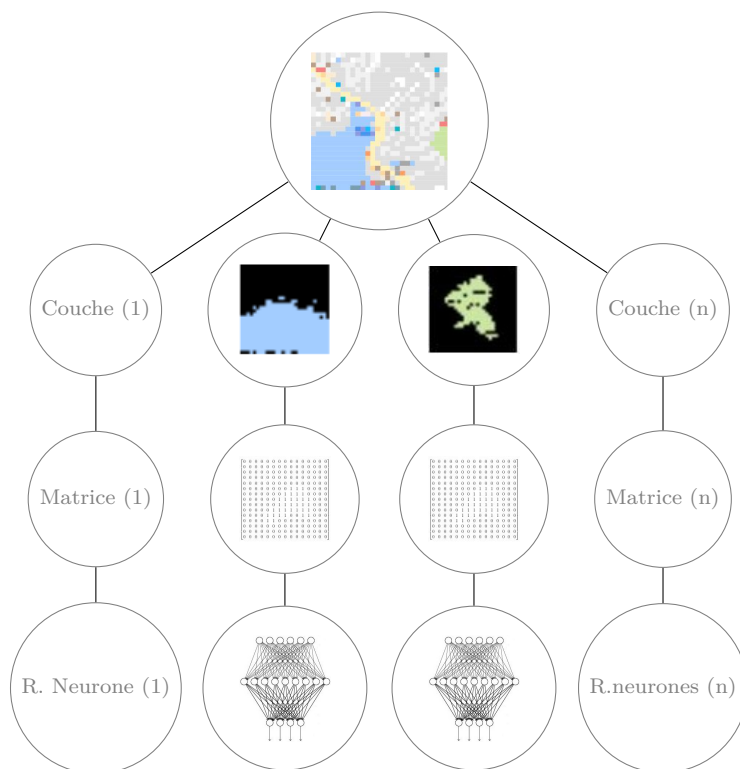
TABLE 1. Exemple des différentes composantes spatiales par couleur  
(\*) : valeur du zoom

Le rôle de chaque réseau de neurones de la figure 11 est de reconnaître la position et la forme de la composante spatiale. L'ensemble des sorties permet la reconstitution d'une portion de l'espace avec une précision meilleure que la méthode basée sur la quantification des couleurs. Néanmoins, cette méthode est plus coûteuse en temps.

## 6. Vers la transcription langagière de l'espace

A terme, au sein d'un système d'aide au déplacement qui permettra aux malvoyants d'être localisé (GPS) et de connaître ainsi la nature et la texture de son environnement proche (cartes et images locales), le rôle de ces réseaux de neurones en série sera de capturer différentes dimensions du paysage dans lequel la personne malvoyante se déplace. Grâce aux classes de textes issus de



FIGURE 11. *Reconnaissance par l'espace par les formes*

la reconnaissance vocale, nous décrivons ce motif spatial en langage maternel ou choisi par la personne malvoyante. Nous nous contentons dans notre test de la transcription simple des noms des composantes spatiales reconnues par le système. Mais il est possible que cette transcription prenne la forme d'un récit plus détaillé. Puisque le système est capable de connaître les coordonnées GPS et d'extraire l'adresse du positionnement de la personne sur la cartographie, le récit vocal peut fournir le nom de la ville ou la rue et la description de cette dernière par rapport au zoom choisi (Ex: *vous êtes situé dans la ville "Avignon", le milieu urbain possède une zone d'intérêt dense, la ville est traversée par un fleuve, on trouve plusieurs espaces verts, etc.*). La description donnée est une description en survol et la finesse de la description dépend du zoom choisi.

Cette approche permettra de reconstituer une description langagière, via la synthèse vocale, de la représentation graphique de l'espace saisie par le malvoyant.

## 7. Conclusion

Dans cet article, nous avons proposé une méthode d'intelligence artificielle pour le traitement visuel de l'information spatiale en mode raster et l'extraction de motifs spatiaux. Les réseaux de neurones utilisés extraient les composantes spatiales de la carte de navigation, à partir des couleurs et de leur interactions spatiales. Leur structure en réseau hyper-connecté et leur capacité d'association et d'apprentissage sont des avantages pour identifier et classer efficacement des motifs spatiaux à partir des couleurs des images. Le prise en compte des différentes échelles d'analyse amène également le réseau de neurones à détecter des objets de différente nature.

Dans le contexte de déplacement des personnes malvoyantes, ces motifs peuvent être appris à la volée et transmis pour information à l'utilisateur en fonction de là où il se trouve, lui permettant ainsi de construire continuellement l'image mentale de son environnement proche. C'est dans cette perspective de transcription langagière de l'espace que ces travaux ont été développés. Des tests restent à réaliser avec des malvoyants pour valider la qualité de la classification et de l'information qui leur sera transmise pour la représentation de cet environnement.

## Remerciements

Mes remerciements à Trent Sartain, développeur senior, pour son blog sur les réseaux de neurones (<https://github.com/trentsartain/Neural-Network>).

## Bibliographie

- Abadi B. P., Martín, Chen J., Chen Z., Davis A., Dean J., Devin M. *et al.* (2016). Tensorflow: A system for large-scale machine learning. *USENIX*.
- Abdi H. (1994). *Les réseaux de neurones*. Sciences et Technologies de la Connaissance. Presss Universitaires de Grenoble.
- Aqil Burney S., Humera T. (2014). K-means cluster analysis for image segmentation. *International Journal of Computer Applications*.
- Bedny M., Richardson R., Hilary. (2015). Visual cortex responds to spoken language in blind children. *JNeurosci - The Journal Of Neuroscience*.
- Bertin J. (1973). *Sémiologie graphique*. Gauthier Villars.
- Bishop C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Boularouk S., Josselin D., Altman E. (2016). Interaction homme machine pour l'orientation spatiale des malvoyants. SAGEO'2016, Nice.
- Celebi E. (2011). Improving the performance of k-means for color quantization. *Cornell University Library NY*.

- Chen T.-W., Chen Y.-L., Chien S.-Y. (2008). Fast image segmentation based on k-means clustering with histograms in hsv color space. *IEEE*.
- Cheng J. X., H, Sun Y., Wang J. L. (2001). Color image segmentation advances prospects. *IEE*.
- Denis M., Gonçalves M.-R. (1995). Mental scanning of visual images generated from verbal descriptions : towards a model of image accuracy. *Neuropsychologia*.
- Doan T.-N., Poulet F. (2014). Large scale image classification: Fast feature extraction, multi-codebook approach and multi-core svm training. *Springer International Publishing*.
- Heaton J. (2015). *Artificial intelligence for Humans - Volume 3: Deep learning and neural networks*. Heaton Research.
- Jodouin J.-F. (1994). *Les réseaux neuromimétiques*. Hermès, Paris.
- Josselin D. (1995). *La déprise agricole en zone de montagne. vers un outil d'aide à la modélisation spatiale couplant systèmes d'induction et d'information géographique*. PhD thesis, Laboratoire de la Montagne Alpine (CNRS URA 344 Grenoble).
- Josselin D., Roussel D., Boularouk S., Saidi A., Matrouf D., Bonin O. *et al.* (2016). Sonorous cartography for sighted and blind people. International Conference on Geographical Information Science AGILE'2016, Helsinki.
- Kubat M., Holte R., Matwin S. (1998). Machine learning for the detection of oil spills in satellite radar images. *Kluwer Academic Publishers, Boston*.
- Morizot A. (2010). La compréhension de l'espace dans les enquêtes de mobilité et de transport. Consulté sur <https://dumas.ccsd.cnrs.fr/dumas-00773396>
- Openshaw S., Openshaw C. (1997). *Artificial intelligence in geography*. Wiley.
- Sammouda R., Adgaba N., Tour A., Al-Ghamd A. (2013). Agriculture satellite image segmentation using a modified artificial hopfield neural network. *Computers in Human Behavior*.
- Sarzeaud O. (1995). *Les réseaux de neurones. contribution à une théorie*. Ouest Editions.
- Zee E. van der, Scholten H. (2013). Application of geographical concepts and spatial technology to the internet of things. *Springer*.